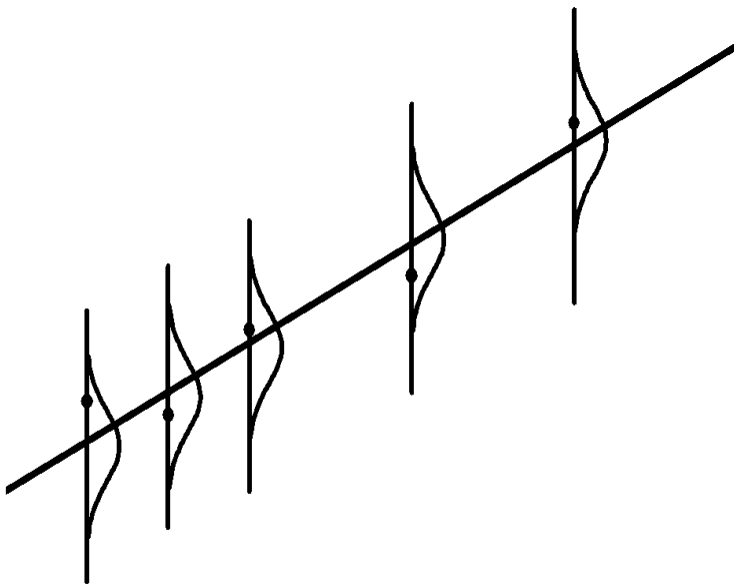
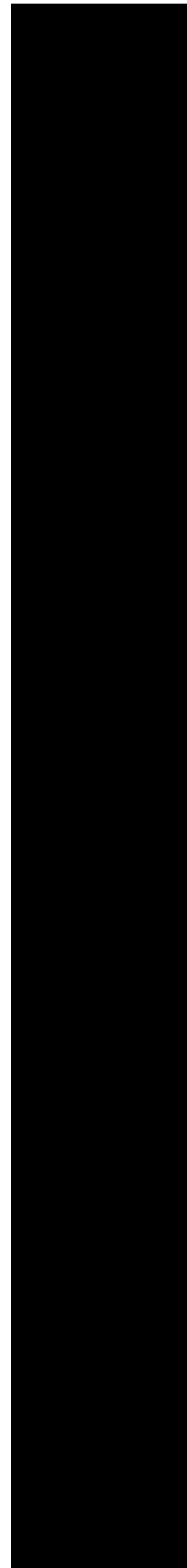




Applying Statistics



**Dan Lurie
Lee Abramson
James Vail**



Statistical distributions

These endpapers provide a convenient summary of most of the distributions that are used in this book. The density (for continuous variables) or the probability function (for discrete variables) is shown along with its mean and variance. The density or probability function is zero outside of the listed range. We also refer to the section in which the distribution is discussed and to the associated quantile table in the appendix, if applicable.

Striving for internal consistency in this book has meant that we have made some changes in the notation used in much of the modern statistical literature. We use y to denote the running variable for each distribution shown. This notation differs from that used in the text for some distributions, where we use a more conventional notation. Specifically, we use Z , t , and χ^2 to denote the standard normal, Student's t , and the chi-square distribution, respectively.

The distributions in these endpapers are shown in the order they are presented in the book.

Continuous distributions		
Distribution	Density	Mean and variance
Uniform (Section 7.2)	$f(y) = \frac{1}{b-a}, a \leq y \leq b$	mean = $(b + a)/2$ variance = $(b - a)^2/12$
Standard uniform (Section 7.3)	$f(y) = 1, 0 \leq y \leq 1$	mean = 0.5 variance = 1/12
Normal (Section 7.4)	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}, -\infty < y < \infty$	mean = μ variance = σ^2
Standard normal (Section 7.5) Table T-1	$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, -\infty < y < \infty$	mean = 0 variance = 1
Lognormal (Section 7.9)	$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(y) - \mu]^2}$ $0 < y < \infty, -\infty < \mu < \infty, \sigma^2 > 0$	mean = $e^{\mu + \frac{\sigma^2}{2}}$ variance = $e^{2\mu + 2\sigma^2} e^{\sigma^2} - 1$
Chi-square (Section 7.10), Table T-2	$f(y) = \frac{e^{-y/2} y^{(v-2)/2}}{2^{(v/2)} \Gamma(v/2)}$ $y > 0, v = 1, 2, 3, \dots$	mean = v variance = $2v$
Student's <i>t</i> (Section 7.11) Table T-3	$f(y) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{y^2}{v}\right)^{-\frac{v+1}{2}}$ $-\infty < y < \infty$	mean = 0 $v = 2, 3, \dots$ variance = $\frac{v}{v-2}$, $v = 3, 4, \dots$
<i>F</i> (Section 7.12) Table T-4	$f(y) = \frac{\left(\frac{v_1}{v_2}\right)^{v_1/2}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} y^{(v_1-2)/2} \left(1 + \frac{v_1}{v_2} y\right)^{-\left(\frac{v_1+v_2}{2}\right)}$ $y > 0$	mean = $\frac{v_2}{v_2-2}$ $v_2 = 3, 4, \dots$ variance = $\frac{2v_2^2}{v_1} \frac{v_1 + v_2 - 2}{v_2 - 4}$ $v_2 = 5, 6, \dots$

(Continued on back flyleaf)



NUREG-1475, Revision 1

Applying Statistics

Manuscript Completed: March 2011

Date Published: March 2011

Prepared by

D. Lurie (Retired)

L. Abramson (Retired)

J. Vail

Office of Nuclear Reactor Regulation

NOTICE

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this book, or represents that its use by such third party would not infringe privately owned rights.

Available as NUREG/CR-1475, Rev 1 from
The Superintendent of Documents
U.S. Government Printing Office
P.O. Box 37082
Washington, DC 20013-7082

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

Preface

The function of the Nuclear Regulatory Commission (NRC) is to regulate the nation's nuclear power industry and the civilian use of special nuclear materials. Its mission is to ensure adequate protection of public health and safety, to promote the common defense and security, and to protect the environment.

Activities in support of the NRC mission include licensing, monitoring, and research. Data and statistical issues arise in all these activities, so that using data and statistical analysis appropriately is an important part of the NRC mission. The goal of this book is to provide the NRC staff with a reference and text on statistical concepts and methods that meet NRC's needs.

The authors include two statisticians and a nuclear engineer who is experienced as a probabilistic risk analyst. The statisticians have a combined total of nearly 70 years of experience at the NRC in applying statistical concepts and techniques to a wide variety of issues in nuclear regulation. These include analysis of licensee data submissions and probabilistic risk assessment of nuclear plants. Although the authors have retired or are about to retire from the NRC, we continue to follow many of the statistical and probabilistic issues that concern the NRC.

We intend this second edition of *Applying Statistics* to be both a reference book on statistical methods and a textbook that may be used in a classroom for teaching this material. While the first edition of this book was written to be used primarily for the licensing and regulation of nuclear power plants, the scope of this edition has been expanded to accommodate the wider variety of applications within the NRC.

Knowledge of statistics can prove invaluable to anyone facing the tasks of organizing and displaying information, formulating quantitative problems, analyzing data, interpreting findings, and evaluating risks. We believe this book can provide that knowledge to readers with a grasp of college algebra. Much of the material can be understood by readers who have a basic background in mathematics..

The title, *Applying Statistics* succinctly expresses our rationale for preparing this material and presenting it in this fashion. To effectively use *Applying Statistics* in your work, it is not necessary to remember on-command results such as formulas for regression lines or procedures for testing hypotheses. Rather, you should be able to recall the essentials of the relevant methods and know where to refresh your knowledge of them.

The presented material is divided into segments, each of which explores a particular aspect of statistics. These segments are organized into chapters, each examining a set of linked statistical ideas. This encourages a pick-and-choose approach, and it allows the presentation of specific illustrative information. Many of the segments contain examples of data that were encountered in the application of statistics at the NRC. Because statistical ideas connect across all of our lives and interests, we also include a number of examples that are not directly related to nuclear regulation.

Chapters 1 through 6 introduce some fundamental statistical concepts and methods. They include definitions, terminology, graphics, probability, and descriptive statistics. These concepts and methods give the reader the beginnings of the vocabulary of statistics and demonstrate how much can be done with just a few basic statistical techniques.

Chapters 7 and 8 present a number of basic continuous and discrete distributions. These functions, being the foundation of statistical estimation and inference, are referenced frequently throughout the book and are used by the NRC in a variety of applications.

Chapters 9 and 10 introduce the main foci of statistics: estimation and inference. Here we explain the philosophy of these concepts and illustrate them with practical examples.

Chapters 11 and 12 present methods for examining data before proceeding with a data analysis. These methods aim to verify assumptions about the structure of the data that justify the use of specific analysis procedures.

Chapters 13 through 19 present methods for drawing inferences from data from continuous distributions. These chapters show how to test statistical hypotheses about data structure for various experimental designs.

Chapter 20 describes the Bayesian methodology and provides examples of its use in risk assessments.

Chapters 21, 22, and 23 focus on three discrete distributions—hypergeometric, binomial, and Poisson—and illustrate their application.

Each of the remaining chapters addresses a specific topic. Chapter 24 focuses on quality control and quality assurance monitoring and testing. Chapter 25 presents a number of nonparametric tests that are used when the distribution from which a sample is drawn is not known. Chapter 26 shows how to identify and deal with outlying observations that may distort statistical inferences from a sample. Chapter 27 introduces the topic of simulation, which is a powerful tool for modeling and analyzing random events.

No discipline is without its jargon—its peculiar, sometimes colorful, phraseology. As they are introduced, the special statistical terms and concepts used in this book are fully defined and explained. They are also highlighted and indexed in the left margin of the text for easy reference.

This publication includes an extensive bibliography that you may find useful in itself, independent of the text. Many of the bibliographic entries are referenced in this book. Some of those entries are general sources whose titles are sufficiently descriptive to help you decide whether you want to explore them further.

This edition of *Applying Statistics* is written under the auspices of the NRC, which recognized the need for this publication and patiently supported our work. We hope this book will meet NRC's expectations and that it will prove valuable in the many activities that require data analysis.

Finally, with respect to errors—of any sort whatsoever—that may still be present in these pages, we are of one mind: They are the result of something that another one of us did or did not do. No one else can share them.

DL
LRA
JAV

March 2011

Acknowledgements

This book is the result of a joint effort which has had the support and collaboration of many individuals both inside and outside the Nuclear Regulatory Commission. We have many people to thank, and hope that we have mentioned all of them.

Professionally, we stand on the shoulders of giants. To practice a daily profession whose foundation was laid by such geniuses as Pascal, the Bernoullis, Gauss, Maxwell, Boltzmann, Galton, Fisher, the Pearsons, Hartley, and Neyman is to be constantly reminded that the vein of statistics runs through all sciences.

Our sincere thanks go to Reginald (Reggie) Mitchell and, Leslie Barnett of the Office of the Chief Financial Officer, who offered us the opportunity and support to pursue this endeavor.

Anthony (Tony) Attard and George (Ed) Powers shared our vision and encouraged us to undertake this important work. They also offered technical support and review throughout the development of this book.

Special thanks go to Robert Easterling of Cedar Crest, NM, who provided a critical review and constructive suggestions for the early chapters. Special thanks also go to Dale Rasmuson, formerly of the Office of Research at the NRC, who helped us in the planning stage of the book.

We—and the book—also benefitted beyond measure from Helen Chang, Keith Azariah-Kribbs and Guy Beltz of the NRC's Division of Administrative Services. Helen Romm of SCA Inc. was helpful throughout the period of editing, formatting and publication. We also acknowledge the support of Mark Cunningham and Lois James both of the Division of Reactor Analysis. Without their support and approval this project would not have proceeded to completion.

Our wives, Ellen Kendrick Lurie, Sondra Glaser Abramson, and Lois Loucks Vail, were always at hand during this effort, making meals, proofreading, forcing short breaks, tutoring language use, and solving household issues. All their selfless assistance and patient devotion have been ours throughout the last

few years. Any day now, they'll get the attention from us that they deserve.

On a personal note, Dan Lurie wishes to dedicate this book to the memory of his late wife, Marti, who contributed greatly to the first edition of *Applying Statistics*. Similarly, Lee Abramson wishes to dedicate the book to the memory of his late wife, Frances, who was supportive of this effort, but unfortunately did not live to see its completion.

Our accolades go to Robin Braverman, Art Director at the American Chemical Society, Washington, DC. Robin was gracious, patient and generous with her time in helping us format, paginate, and style the book to bring it to its present shape.

The first edition of *Applying Statistics* was published in 1994 and was coauthored by Roger Moore, formerly Chief of the Applied Statistics Branch of the NRC. Roger's invaluable contributions to the 1994 edition still resonate in this second edition.

Table of Contents

Preface	iii
Acknowledgements	vii
Table of Contents	ix
Chapter 1 Introduction	1
1.1 What to look for in Chapter 1	1
1.2 What is statistics?	2
1.3 Probability and statistics	3
1.4 Data	4
1.5 Scales of measurement.....	5
1.6 Four basic concepts in statistics	7
1.7 Evaluating statistical statements	8
1.8 Misconceptions about statistics	11
1.9 Spreadsheet computation	12
Chapter 2 Descriptive statistics	15
2.1 What to look for in Chapter 2.....	15
2.2 Descriptive statistics.....	16
2.3 Measures of centrality	17
2.4 The weighted mean	21
2.5 Spreadsheet functions for measuring centrality	23
2.6 Measures of dispersion: range and quantiles	24
2.7 Variance and standard deviation	28
2.8 Spreadsheet functions for measuring dispersion.....	29
2.9 Descriptive statistics with a hand-held calculator.....	30
2.10 Descriptive statistics for coded data	31
2.11 Skewness and kurtosis	33
2.12 An empirical rule for a mound-shaped dataset	34
2.13 Estimating the standard deviation from the range	36
2.14 Chebyshev's inequality	37

Chapter 3 Statistical graphics	39
3.1 What to look for in Chapter 3.....	39
3.2 What's in a graph?	40
3.3 The pie chart	41
3.4 Suggestions for constructing pie charts	43
3.5 The bar chart	44
3.6 Suggestions for constructing bar charts	47
3.7 The histogram	48
3.8 Suggestions for constructing histograms	50
3.9 The box plot	51
3.10 Suggestions for constructing box plots	56
3.11 The stem-and-leaf display	57
3.12 Suggestions for constructing a stem-and-leaf display ..	60
Chapter 4 Basics of probability	61
4.1 What to look for in Chapter 4.....	61
4.2 The concept of probability	62
4.3 Sample spaces and events	63
4.4 Basic set theory	64
4.5 Basic rules and principles of probability.....	67
4.6 Marginal and joint probabilities	70
4.7 Conditional probability	73
4.8 Bayes' Theorem	78
4.9 Probability Estimation	79
Chapter 5 Errors	83
5.1 What to look for in Chapter 5.....	83
5.2 Errors, errors, ... everywhere	83
5.3 Characterizing errors: accuracy and precision	84
5.4 Uncertainty	87
Chapter 6 Random variables.....	89
6.1 What to look for in Chapter 6.....	89
6.2 Random variables	90

6.3	Distributions of discrete random variables	91
6.4	Distributions of continuous random variables	95
6.5	Distributions of discrete bivariate random variables	97
6.6	Distributions of continuous bivariate random variables	101
6.7	Expected value and variance	102
6.8	Linear combinations	104
6.9	Linear contrasts	106
6.10	Mean square error	107
Chapter 7 Continuous distributions		109
7.1	What to look for in Chapter 7	109
7.2	Uniform distribution	110
7.3	Standard uniform distribution	112
7.4	Normal distribution	113
7.5	Standard normal distribution	116
7.6	Table lookup for the normal distribution	117
7.7	Spreadsheet functions for normal probabilities	121
7.8	The Central Limit Theorem	123
7.9	Lognormal distribution	126
7.10	Chi-square distribution	129
7.11	Student's <i>t</i> -distribution	132
7.12	<i>F</i> -distribution	135
7.13	Exponential distribution	139
7.14	Gamma distribution	141
7.15	Beta distribution	142
Chapter 8 Discrete distributions		147
8.1	What to look for in Chapter 8	147
8.2	Discrete uniform distribution	148
8.3	Sampling for attributes	150
8.4	Sampling with and without replacement	151
8.5	Bernoulli distribution	152
8.6	Hypergeometric distribution	154
8.7	Binomial distribution	158

8.8	Geometric distribution	161
8.9	Negative binomial distribution	163
8.10	Poisson distribution	165
Chapter 9	Estimation	169
9.1	What to look for in the remainder of this book	169
9.2	What to look for in Chapter 9.....	170
9.3	Estimation and inference	170
9.4	Elements of estimation	171
9.5	Point estimators	172
9.6	Interval estimators	174
9.7	Confidence intervals for a mean.....	176
9.8	Two-sided 95% confidence intervals for a mean	176
9.9	One-sided 95% confidence intervals for a mean	179
9.10	Confidence intervals with an arbitrary confidence level	180
9.11	Confidence intervals for unknown σ	181
9.12	Statistical tolerance limits for a normal population	183
9.13	Confidence intervals for a variance	185
9.14	Sample size determination: σ known	186
9.15	Sample size determination: σ unknown	187
Chapter 10	Inference	191
10.1	What to look for in Chapter 10.....	191
10.2	Testing statistical hypotheses: setting the stage.....	192
10.3	Terminology	194
10.4	Null and alternative hypotheses: examples	198
10.5	Consequences of hypothesis testing	199
10.6	Guilty until found innocent	200
10.7	Finally.....	201
Chapter 11	Goodness-of-fit tests.....	203
11.1	What to look for in Chapter 11.....	203
11.2	Testing goodness-of-fit	204
11.3	Chi-square test for discrete distributions.....	206

11.4	Chi-square test: sample-size considerations	209
11.5	Chi-square test for normality: known parameters	210
11.6	Chi-square test for normality: unknown parameters	212
11.7	Empirical cumulative distribution function	216
11.8	Kolmogorov-Smirnov goodness-of-fit test.....	217
11.9	Shapiro-Wilk (W -) test for normality	221
11.10	D'Agostino (D') test for normality	224
Chapter 12 Contingency Tables.....		229
12.1	What to look for in Chapter 12.....	229
12.2	Contingency tables	230
12.3	Structure of contingency tables	230
12.4	Independence	232
12.5	Testing independence	233
12.6	Special case: 2×2 contingency tables.....	236
12.7	Fisher's exact probability test	238
12.8	Simpson's paradox—better watch out!	241
12.9	A contingency table look-alike: McNemar's test statistic	242
Chapter 13 Tests of statistical hypotheses: One mean.....		247
13.1	What to look for in Chapter 13.....	247
13.2	A test of the mean: σ known	248
13.3	A one-sided test: A different view	249
13.4	Power of a test of hypothesis	251
13.5	Operating characteristic curve.....	255
13.6	More power to you	255
13.7	Testing a mean when σ is unknown	256
13.8	Hypotheses with two-sided alternatives.....	257
13.9	Required sample size to test a mean: σ known	258
13.10	Required sample size to test a mean: σ unknown ...	261
Chapter 14 Tests for statistical hypotheses: Variances.....		263
14.1	What to look for in Chapter 14.....	263
14.2	Why worry about variances?	264

14.3	Testing a single variance.....	265
14.4	Testing equality of two variances.....	266
14.5	Testing homoscedasticity with samples of equal size.....	269
14.6	Pooling variances.....	271
14.7	Testing equality of variances with unequal sample sizes.....	272
Chapter 15 Tests for statistical hypotheses: Two means ..		277
15.1	What to look for in Chapter 15.....	277
15.2	Hypotheses about the means of two populations	278
15.3	Procedure 1: Paired observations	279
15.4	Procedure 2: Variances known	284
15.5	Procedure 3: Variances unknown but assumed equal	286
15.6	Procedure 4: Variances unknown and unequal	288
15.7	An example to summarize the four procedures	290
15.8	Required sample size.....	293
Chapter 16 One-way ANOVA.....		295
16.1	What to look for in Chapter 16.....	295
16.2	One-way ANOVA: Data structure	296
16.3	Descriptive statistics.....	298
16.4	Model and assumptions	299
16.5	Partition of the total sum of squares	302
16.6	The one-way ANOVA table	305
16.7	Constructing an ANOVA Table with Excel.....	307
16.8	Duncan's multiple range test.....	308
16.9	T-test and the ANOVA equivalence	312
Chapter 17 Two-way ANOVA		315
17.1	What to look for in Chapter 17.....	315
17.2	Two-way factorial designs	316
17.3	Randomized complete block design.....	317
17.4	Data structure and model: No replication.....	318

17.5	ANOVA for a two-way factorial design without replication.....	322
17.6	Excel calculations for the ANOVA table: No replication.....	324
17.7	Balanced two-way factorial designs with replication .	325
17.8	Other multifactor designs	330
Chapter 18	Regression.....	331
18.1	What to look for in Chapter 18.....	331
18.2	Concepts and terms from algebra and geometry	332
18.3	The concept of regression.....	333
18.4	Regression models	335
18.5	Simple linear regression.....	337
18.6	Fitting a line to data.....	339
18.7	The method of least squares and the regression line.....	341
18.8	Geometric interpretation of regression components .	345
18.9	Partition of the total sum of squares	346
18.10	Regression ANOVA for a single independent variable.....	348
18.11	Using Excel to construct the regression line	350
18.12	Using Excel for regression analysis.....	353
18.13	Hypothesis testing and confidence intervals for the slope	357
18.14	Hypothesis testing and confidence intervals for the intercept	359
18.15	Regression through the origin	361
18.16	Multiple linear regression	364
18.17	Prediction	366
Chapter 19	Simple linear correlation.....	369
19.1	What to look for in Chapter 19.....	369
19.2	Basics of Simple linear correlation.....	370
19.3	The correlation coefficient	372
19.4	Excel's routines for calculating correlation.....	376

19.5	Testing the correlation coefficient	377
19.6	Confidence interval for the correlation coefficient	380
19.7	Testing equality of two correlation coefficients.....	383
19.8	Comparison of regression and correlation analyses .	384
Chapter 20 Bayesian probability inference		387
20.1	What to look for in Chapter 20.....	387
20.2	Motivation for Bayesian inference	387
20.3	Bayesian inference.....	388
20.4	Bayesian parameter estimation	390
20.5	Conjugate distributions.....	391
20.6	Noninformative prior distributions	396
20.7	Nonconjugate prior distributions	397
20.8	Bayesian hypothesis testing	398
20.9	Bayes factors	399
20.10	Consistency of the prior with the observed data	400
Chapter 21 Hypergeometric experiments		403
21.1	What to look for in Chapter 21	403
21.2	Basics of the hypergeometric distribution	404
21.3	Estimates of the proportion and number of successes	406
21.4	Tests of hypothesis	406
21.5	Sample size considerations	408
21.6	Normal approximation to the hypergeometric distribution.....	409
Chapter 22 Binomial experiments		413
22.1	What to look for in Chapter 22	413
22.2	Prerequisites for a binomial experiment.....	414
22.3	Binomial probabilities	415
22.4	Examples of binomial experiments	416
22.5	Mean and variance of a binomial variable	420
22.6	Normal approximation to the binomial	422

22.7	Confidence interval for proportion: Normal approximation applies	423
22.8	Confidence interval for proportion: Normal approximation does not apply.....	424
22.9	Confidence interval for a sample with no defects.....	426
22.10	Sample size for political polling	427
22.11	Binomial approximation to hypergeometric distribution.....	429
Chapter 23 Poisson experiments.....		431
23.1	What to look for in Chapter 23	431
23.2	Prerequisites for a Poisson experiment	432
23.3	Poisson probabilities	433
23.4	Applications.....	435
23.5	Parameter testing and confidence intervals.....	436
23.6	Approximations to the binomial distribution	438
23.7	The normal approximation to the Poisson distribution	439
Chapter 24 Quality assurance.....		441
24.1	What to look for in Chapter 24.....	441
24.2	The concept of quality assurance	442
24.3	Process control: Building in quality	443
24.4	Control charts for means	444
24.5	Run test for control charts	446
24.6	Control charts for variability	448
24.7	Control charts for attributes	451
24.8	Acceptance sampling: Verifying quality	451
24.9	The A/Q criterion: Rules of the game	454
24.10	Probability calculations for the 95/95 criterion.....	456
24.11	What's wrong with this picture?	458
Chapter 25 Nonparametric statistics.....		461
25.1	What to look for in Chapter 25.....	461
25.2	Nonparametric methods	462

25.3	Test of randomness: The runs test	463
25.4	Test for location: The sign test	466
25.5	Test for location: Wilcoxon signed ranks test.....	469
25.6	Test of locations with two matched samples: Sign test	472
25.7	Test of locations with two samples: Wilcoxon matched pairs test	474
25.8	Test of locations; two independent samples: Wilcoxon rank sum test	475
25.9	Test of locations with several samples: The median test.....	478
25.10	Test of locations with several samples: The Kruskal-Wallis test.....	482
25.11	Test of location for related samples: The rank ANOVA test.....	486
25.12	Test of location for related samples: The Friedman test.....	489
25.13	Test of variances with two samples: Squared ranks test	491
25.14	Test of variances with several samples: The k -sample squared ranks test.....	494
25.15	Spearman's test of independence for two populations.....	498
Chapter 26 Outliers		503
26.1	What to look for in Chapter 26.....	503
26.2	What is an outlier?	504
26.3	Box plot procedure for outlier identification.....	505
26.4	Dixon's procedure for outlier identification	507
26.5	Grubbs' tests for outliers	508
26.6	Considerations in outlier rejection.....	511
Chapter 27 Simulation		513
27.1	What to look for in Chapter 27.....	513
27.2	Introduction to simulation	513

27.3	Generation of random numbers.....	515
27.4	Estimation of definite integrals	517
27.5	Generation of normal variates	520
27.6	Generation of arbitrary variates	521
27.7	Confirmation of the central limit theorem	523
Appendix / Tables		527
Bibliography.....		585
Index.....		601

1

Introduction

1.1 What to look for in Chapter 1

Chapter 1 introduces the discipline and the practice of statistics in commonly understood terms, thereby setting the stage for the ensuing chapters. Along the way, you may find that some accepted statistical conceptions are, in fact, statistical misconceptions. You will soon be equipped with a straightforward set of criteria you can use to appraise many statistical notions or statements. Among the fundamental statistical concepts developed and illustrated in this chapter—and expanded upon in the rest of the book—are the following:

- *statistics (the discipline), §1.2*
- *deductive and inductive reasoning, §1.3*
- *qualitative and quantitative information, §1.4*
- *discrete and continuous variables, §1.4*
- *scales of measurement, §1.5*
 - nominal (categorical) scale*
 - ordinal (relative) scale*
 - interval (metric, continuous) scale*
 - ratio scale*
- *population, §1.6*

- *sample*, §1.6
- *parameter*, §1.6
- *statistic (a summary of data)*, §1.6
- *Huff's criteria*, §1.7

Section 1.7 also provides a set of five criteria—called Huff's criteria—to be used in evaluating statistical statements. Although these criteria are intuitive, we sometimes fail to check whether they are met. The chapter presents examples where one or more of Huff's criteria are violated.

Section 1.9 concludes with general guidelines for the use of Excel®'s spreadsheet functions and routines in carrying out statistical calculations and analyses. Excel is the only spreadsheet discussed in this text as it is the official spreadsheet program used at the U.S. Nuclear Regulatory Commission (NRC).

Although Excel can perform most common statistical analyses, it is not a statistical package. Several statistical packages on the market are capable of many specific and general tasks, but this book does not discuss any of these packages. The interested reader who requires a specific or general package may refer to any of the many articles available on the Internet. A readily available source that reviews most of these packages may be found on the Internet. Look for "Comparison of Statistical Packages" in Wikipedia.

1.2 What is statistics?

One can approach an unfamiliar or little-understood subject in many ways. One useful technique is to settle on some definitions of the basic elements of the subject and then build on those definitions as additional topics and details arise. This will allow you to develop enough knowledge and comfort to form a basis for learning about the subject in more depth.

statistics There is no universally accepted definition of *statistics*. Many definitions or characterizations of statistics have been suggested, some of which are listed below:

- A science of collecting, interpreting, and representing data
- An art and science of treating data
- A science that deals with the collection of limited data to form appropriate conclusions about the general case
- A language, a mechanism for creating and communicating quantitative concepts and ideas
- A science of decision making in the face of uncertainty
- A mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data

As this list suggests, statistics is a discipline with many facets. However, for the purpose of this book, there is no need for a formal definition of statistics. Rather, we will use statistics as a generic term that refers to whatever topic is being considered at the time.

1.3 Probability and statistics

*deductive
reasoning*

*inductive
reasoning*

Statistics is often considered a branch of mathematics, and so is probability.¹ However, whereas the subject of probability is based on a formal development and rigorous proofs, the practice of statistics also involves empirical information that may or may not represent the entire story. Probability uses *deductive reasoning* in its development, whereas statistics uses primarily *inductive reasoning*. Probability starts with mathematical assumptions and deduces the probabilities of specific outcomes. Statistics begins with data and works backwards to infer what sort of probabilistic structure could have given rise to the observed data. The following examples illustrate the difference between deductive and inductive reasoning.

Suppose that a jar contains 10 red marbles, 20 white marbles, and 30 black well-mixed marbles. Blindfolded, we reach into the jar and remove one marble. Under the assumption that every marble has the same probability of being selected, we can calculate the probability that the marble we picked is, say, red. Similarly, we may remove five marbles and answer questions such as what is the probability that the marbles are all of the same color, that there is exactly one red marble, that there are at least two white marbles, or that there are more black marbles than any other color. The answers to these questions are all examples of deductive reasoning, which in this case, can be described as making statements about the unknown contents of our closed hand based on the known contents of the jar.

Contrast this scenario with the situation where we draw five marbles from a jar containing unknown quantities of red, white, and black marbles. We open our hand and see that we have three red and two white marbles. Now we can answer questions about the contents of the jar, such as what is the ratio of red to white marbles or what is the proportion of black marbles in the jar, even though we did not see a single black marble in our hand. This is an example of inductive reasoning, which in this case can be described as making statements about the unknown contents of the jar based on the known contents of our hand.

¹ Probability is defined and discussed in Chapter 4.

In summary, probability makes inferences about a potential sample² from a known population using deductive reasoning. Statistics makes inferences about an unknown population from a known sample using inductive reasoning.

1.4 Data

Statistical investigations are based on data. Understanding processes, products, and issues generally involves identifying a pertinent set of items and measuring or recording various characteristics of these items. These recorded characteristics, or variables, are the data from which we try to extract information. For example, the items of interest could be nuclear power plants, on which the NRC collects much data pertaining to operational reliability. The investigation may address issues such as how reliability changes with age and differs among manufacturers or designs.

A characteristic can be classified in one of two ways:

<i>quantitative</i>	<i>quantitative</i> , when it can be assigned a numeric value (such as 0, 1, 7.5, $\sqrt{17}$, or $-1/4$)
<i>qualitative</i>	<i>qualitative</i> , when it cannot be assigned a compelling numeric value (such as name, birthplace, or gender)
<i>variable</i>	Be it quantitative or qualitative, a characteristic is called a <i>variable</i> because its specific value or nature is not known before the item is examined. We determine a <i>value</i> for a characteristic or variable by making a <i>measurement</i> of it (e.g., how big is it? how strong is it?) with an instrument or by making an <i>observation</i> of it (e.g., what color is it? what kind is it?) by examining it. The value for a characteristic of a specific item is called a <i>datum</i> .
<i>value</i>	
<i>measurement</i>	
<i>observation</i>	
<i>datum</i>	

Generally, when gathering information, we will examine more than one item and, for each characteristic of interest, determine its value. The reason we examine more than one item is that, almost always, items identical in name and in definition are not necessarily—or even probably—identical with respect to all of their characteristics. Now we have *data* (the plural of datum) to work with, and the door to statistical ideas and procedures is open.

data

Working with and discussing several items and several variables can be difficult in many settings and with many

² The terms “sample” and “population” are defined in Section 1.6.

<i>dataset</i>	audiences. A useful device, with both theoretical and practical benefits, is the <i>dataset</i> , literally a set of data. Datasets are commonly displayed in two-way tables, with the rows representing the items and the columns representing the variables.
<i>record field cell value</i>	Modern database management theory and statistical computing practice strongly influence this dataset structure. In these disciplines, each row of the table is a <i>record</i> , each column is a <i>field</i> , each entry at a row-column intersection (often called a <i>cell</i>) is a <i>value</i> , or <i>observation</i> , and the entire table is a <i>file</i> .
<i>observation file discrete variable</i>	A variable is said to be <i>discrete</i> if it can assume only a finite or countably infinite number of values. Examples of discrete variables include the number of defective units in a batch of manufactured items and the number of atomic disintegrations in a sample of radioactive material in a given unit of time.
<i>continuous variable</i>	In contrast, a variable is said to be <i>continuous</i> if it can assume any value within its range. Examples of continuous variables include linear dimensions, such as weight, height, volume, and temperature, and wavelengths associated with the color spectrum.

Quantitative variables can be either discrete or continuous, while qualitative variables must be discrete. Because the mathematics of working with continuous variables is usually easier than that for discrete variables, it is often preferable to treat discrete variables as continuous. Examples are the cost of printing a page of text on a modern printer or the cost of fueling your car. Consequently, we recommend treating them as continuous, unless there are compelling reasons to the contrary.

1.5 Scales of measurement

scales of measurement Variables (i.e., measured values or observations) can be categorized as arising from one of four *scales of measurement*, beginning with a nominal scale. The other three scales, called ordinal, interval, and ratio, are successively built from a nominal scale by adding certain conditions. Without knowing the scale used in making the observations, we risk producing erroneous analyses because each scale's defining conditions determine which set of tools is appropriate for the analysis.

- nominal scale** A **nominal scale** (sometimes called a **categorical scale**) employs words, symbols, or numbers to identify the categories or groups to which the items of interest belong. Thus, a nominal scale implies no special sequencing or ordering; that is, the {yes, no} scale carries no more information than the {no, yes} scale. Similarly, the {Democrat, Independent, Republican} scale carries no more information than the {Republican, Democrat, Independent} scale. A nominal scale applies only to qualitative variables.
- categorical scale**
- ordinal scale, relative scale** An **ordinal scale** (sometimes called a **relative scale**) adds an ordering or ranking of the characteristics designated by a nominal scale. Thus, an ordinal scale captures additional information, such as when ordering vehicles by their weight or horsepower (e.g., {truck, auto, motorcycle}). An ordinal scale also applies to qualitative variables. For example, ordering employees by their highest degree (i.e., bachelor, master, or Ph.D.).
- interval scale** An **interval scale** (sometimes called a **metric scale** or **continuous scale**) preserves the characteristics of an ordinal scale and, in addition, assigns a value or numerical “distance” between each pair of objects. Thus, the differences between values are important in an interval scale. For example, an interval scale that measures distances between pairs of cities permits the determination of which two cities are closest, which two are farthest apart, and which pairs of cities are the same distance apart. An interval scale applies only to quantitative variables.
- metric scale**
- continuous scale**
- ratio scale** A **ratio scale** is an interval scale with a physically meaningful and definable zero value, which is the value reported when a characteristic is not detected by the measuring process. An important property of a ratio scale is that the ratio of any two measured values of the same type is independent of the unit of measurement (e.g., the ratio of the width to the length of a rectangle is independent of whether both measurements are made in inches, feet, kilometers, or furlongs). Some other examples of ratio scales are weight (kilograms or pounds), length (meters or inches), and volume (gallons or liters); the ratios of all of these are independent of whether metric, English, or other units are used. In contrast, temperature, an interval-scale variable, is not a ratio-scale variable. For example, the ratio of 50 degrees Celsius (°C) to 40 °C is 1.2, while the ratio of their Fahrenheit (°F) equivalents, 122 °F and 77 °F, is 1.58. Because it is also an interval scale, a ratio scale applies only to quantitative variables.

1.6 Four basic concepts in statistics

Four concepts form the basis for statistical procedures and inference. Consequently, it is essential to address them at the beginning of any statistical investigation.

population A **population** is a collection of measurements made on items defined by some characteristic of the items. To be exact, an observation (such as weight or activity status) is associated with each item. (There can also be more than one observation for each item, usually written as a vector.) The collection of all of these items' characteristics makes up the population. Examples are the population of 20,000 pellet enrichments (e.g., the set {85.4%, 86.1%, 86.0%, ...}) and the population of the ages of the employees in our organization (e.g., the set {37, 74, 66, 62, ...}).

sample A **sample** is a subset of the population containing one or more items. If the sample contains all the items in the population, that sample is termed a **census**.

parameter A **parameter** is a numerical characteristic of the population; its value is a function of the values of the variables of the items in the population. Examples of parameters are the average of all the hourly wages in the population and the percentage of defective items in the population.

statistic A **statistic** (note the singular form!) is a numerical measure of a sample; its value is a function of the values of the variables of the items in the sample. Thus, the sample average is a statistic. The ratio of the number of defective items in the sample to the number of items in the sample is another statistic.

- A population may be a collection of measurements (or responses) rather than the individual physical items (e.g., automobiles, houses, employees) that were examined to obtain the measurements (or responses).

universe In some contexts, the terms **universe** or **parent population** are used instead of population. The term parent population serves as a reminder that a sample is an "offspring" of the population of interest.

- A population's size (i.e., the number of its elements) can be finite or infinite. A population can, and usually does, have more than one associated parameter.

- A sample can, and usually does, have more than one statistic.

To make valid statistical inferences about a population from a sample, the sample must be random. This will be discussed later in Chapter 6.

Example 1.1, in which the air mass leakage rate from a pressurized containment is determined, illustrates the four concepts of this section.

Example 1.1. Containment integrated leakage rate testing. In determining whether the leakage rate from a pressurized containment is acceptable, the licensees conduct periodic tests. The leakage rate is determined by changes in containment air mass over time. At any specific time, air mass is determined by the spatial distribution of the temperature, pressure, and relative humidity within the containment. By use of sensors carefully placed within the containment, these three distributions are concurrently sampled. The relevant sensor statistics are the average temperature, pressure, and relative humidity of the respective samples. The population parameters that are estimated from the samples are the average temperature, pressure, and humidity over the entire containment air mass.

1.7 Evaluating statistical statements

To evaluate any statistical statement, it is necessary to ascertain the relevance and validity of both the data and the analysis on which the statement is based. It's not necessary to be a statistician to react sensibly to a statistical statement or analysis. Almost all problems attributed to "bad" data or to "bad" statistics are avoidable. We simply must be aware of pitfalls along the way and take proper precautions in dealing with them.

Huff's criteria The following five questions, adapted from Huff (1954) and hereafter called **Huff's criteria**, provide a first line of defense against an incorrect interpretation of any statistical statement:

- Who says so? (*Does he/she have an axe to grind?*)
- How does he/she know? (*Does he/she have the resources to know the facts?*)
- What's missing? (*Does he/she give us a complete picture?*)
- Did someone change the subject? (*Does he/she offer us the right answer to the wrong problem?*)
- Does it make sense? (*Is his/her conclusion logical and consistent with what we already know?*)

Throughout these discussions and, even more importantly, long afterwards, as we go about our regular business, personal, or professional activities, it is important to maintain a healthy skepticism about statistical claims and

counterclaims. Try to evaluate statistical statements with Huff's criteria in mind. Indeed, we will find them applicable to nonstatistical statements as well, adding vigor and rigor to our analyses and value and meaning to our conclusions.

Note that the first of Huff's criteria is the only one that does not involve the content of the statement; rather, it deals with motivation. It is important to note that any incorrect interpretation of a statement because one or more of the other criteria implied by Huff's questions is violated does not depend on whether the violation is deliberate. While motivation might certainly be a factor in understanding why Huff's criteria are violated, it is irrelevant in assessing their applicability.

Example 1.2. Evaluating the regulator. A group of utilities surveyed opinions on the efficiency, fairness, and impact of the agency that regulates them. The associated report was highly critical and stated that the great majority of the people interviewed agreed that the agency was detrimental to the safe and efficient operation of their facilities.

Sounds bad, doesn't it? The question is, "Who said so?" It turns out that the participants in the survey were all employees of the utilities involved. Those participants surely have their own axes to grind; they did not invite the regulator to oversee them, and they would be more flexible in their operations without someone looking over their shoulder. Whether they agree or whether their intentions are good, their opinion is biased. Therefore, the survey violated the first of Huff's criteria, "Who says so?" It also violated Huff's third criterion, because it reflected only the utilities' perspective and was therefore incomplete.

Example 1.3. Challenger's O-ring failure. You may remember the cold morning of January 28, 1986, when the space shuttle Challenger broke apart in mid-air. The cause of the accident was the failure of one or more of the six O-rings in the shuttle to seat properly, resulting in a burn through the external gas tank. A chart similar to Figure 1.1 was instrumental in convincing managers from the National Aeronautics and Space Administration and their contractors that it was safe to launch despite a predicted launch temperature of about 32 °F (much lower than any previous launch) on the grounds that the launch temperature does not have an effect on the integrity of the O-rings.

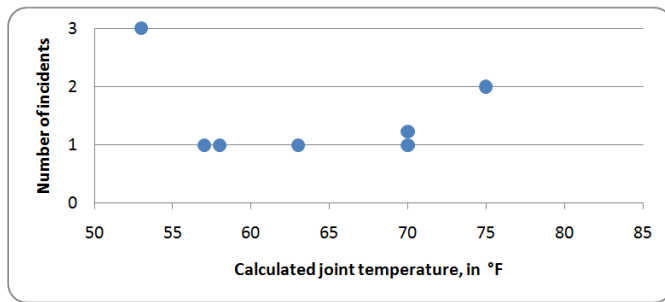


Figure 1.1. Distressed O-rings vs. temperature, 7 launches

scatter plot Figure 1.1 is a variation of a graph called a *scatter plot*. It plots the number of incidents of O-ring degradation as a function of temperature. Even though the highest number of incidents (3) occurred at the lowest temperature, the second highest number of incidents (2) occurred at the highest temperature. Based on the absence of any obvious pattern, it is plausible to conclude from the data in Figure 1.1 that temperature does not affect O-ring performance.

However, Figure 1.1 is based only on the seven previous launches where there were one or more incidents of O-ring degradation. In fact, there were a total of 24 previous space shuttle launches, of which 17 had no such incidents. Clearly, this analysis violates the third of Huff's criteria. Crucial information is missing in Figure 1.1!

Figure 1.2 provides the full picture of the relation between temperature and O-ring performance. It shows that there were 17 launches, all conducted at joint temperatures above 65 °F, without O-ring distress. When added to the previous launches from Figure 1.1, the data clearly indicate that the likelihood of O-ring distress increases as the temperature decreases below 65 °F. So what went wrong with the Challenger? The launch temperature was 31 °F. A clear extrapolation of the given data implies a significant increase in the likelihood of O-ring distress over previous launches (the lowest previous temperature was 53 °F). Although we cannot be certain that managers would have changed the decision to launch on January 28 if the data in Figure 1.2 had been available, this incident is a case where appropriate statistical analysis may literally have been a matter of life and death.

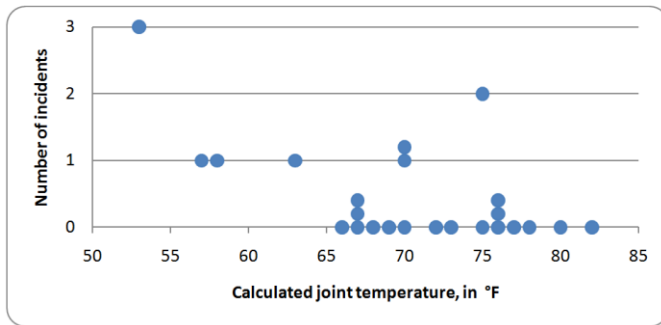


Figure 1.2. Distressed O-rings vs. temperature, 24 launches

A detailed discussion of the analysis of the Challenger disaster is given in Dalal (1989).

Example 1.4. Stolen cars. A newspaper reported that the top five car models stolen over 5 consecutive years in the State were all Toyota Camrys. Based on this information, some readers decided not to buy a Camry. Their interpretation of the report may be faulty. Could it be that more Camrys are stolen simply because more Camrys are on the road? Or could it be that they are more desirable? And if more desirable cars are likely to be stolen, why isn't Ferrari on the list? What's missing is, at the very least, the number of each model on the road and the number of each model that is stolen. Without this information, Huff's third and fourth criteria are violated.

Example 1.5. Drunk drivers. It was reported that drunk drivers caused 15% of the fatal accidents on the road. This means that sober drivers caused 85% of the accidents. Should we then conclude that sober drivers should not be allowed to drive?

This example clearly illustrates the violation of Huff's third and fifth criteria. What is missing is the proportion of fatal accidents involving drunk drivers to the number of drunk drivers on the road and a similar proportion for sober drivers. One should also ask whether the conclusion makes sense or whether it conflicts with common knowledge.

1.8 Misconceptions about statistics

The credibility and quality of statistical investigations depend on the data and the analysis. Deliberate or inadvertent biases can adversely affect an investigation. Except for the last one, the following book titles (all of which are real) reflect some of the many misconceptions and biases held by many people about statistics (and about statisticians):

Use and Abuse of Statistics (Reichmann, 1971)

How to Lie with Statistics (Huff, 1954)

Flaws and Fallacies in Statistical Thinking (Campbell, 1974)

How to Tell the Liars from the Statisticians (Hooke, 1983)

The Tyranny of Numbers (Eberstadt, 1995)

Of course, no discourse on statistics escapes the canard “there are three kinds of lies—lies, damned lies, and statistics.”³ We really don’t have to look very hard or very far to find such “wisdom” and/or gobbledygook. All we have to do is to pay attention to the news media, read a few technical journals, listen to our doctors and dentists, take note of our politicians, and participate in everyday life. The fact is that everyone is well acquainted with situations in which statistics are used and misused. The real need is to develop and improve the insights and the tools to sort through the blizzard of quantitative “facts” that surrounds us. Indeed, our load is lightened considerably if we never lose sight of Mosteller’s insightful rejoinder, “It’s easy to lie with statistics. But it is easier to lie without them.”⁴

1.9 Spreadsheet computation

Displaying data and extracting and communicating information from data in a convincing manner require data plots and computational analyses. Having software to do this work is very useful, and much software is available. Many basic plots and analyses can be done with Excel, which is the official spreadsheet of the NRC. We will illustrate and provide guidance on the use of Excel throughout this book.

Statistical functions often operate on a selected set of values. The collection of those values must be identified to the function by selecting (either listing or highlighting) the cells that contain these values. In Excel, for data listed in a single column or a rectangular array, the upper left-most cell and the lower right-most cell (separated by one or two periods) must be listed (e.g., A1..B100), or highlighted by dragging the mouse from cell A1 to cell B100. Data that are not in a column or a rectangular array can be entered as multiple segments separated by commas, such as A1..B25,C7..F35. For example, the average function is written as =AVERAGE(A1..A100) or =AVERAGE(A1..B25,C7..F35).

- Excel requires an equal sign as a prefix in all functions.

³ Attributed by Mark Twain (Samuel Clemens, 1835–1910) to the British Prime Minister Benjamin Disraeli (1804–1881), Source: *The Pocket Book of Quotations*, edited by Henry Davidoff, Pocket Books, Inc., New York, NY, 1952.

⁴ Frederick Mosteller, quoted in *Chance*, Vol. 6, No. 1, 1993, pp. 6–7.

- Excel uses a colon (:) to separate cells, such as A1:B25. However, if a period, two periods, or three periods are used as separators (such as A1.B25, A1..B25, or A1...B25), Excel accepts them but automatically converts these separators to single colons.

To see a hypertext listing of the statistical functions available from within Excel, click on Help—Table of contents—Working with data—Function reference—Statistical functions. Click on any of these functions, and a help panel will drop down with an explanation and an example of the function and its use.

In some instances, Excel offers complete routines that produce a set of relevant calculations. The use of such routines requires that Excel's Data Analysis Tools (sometimes called Data Analysis Pack) be installed first with Excel's Options Add-ins. Calls to routines are made by clicking on Data—Data Analysis—routine name and then following the instructions on the screen. Help and instruction are also available to execute the routine. The use of such routines is explained and illustrated in this book in the context of their applications.

2

Descriptive statistics

2.1 What to look for in Chapter 2

Studies and investigations often involve datasets of various sizes. To communicate essential and informative characteristics of such datasets, we use descriptive statistics. Properly presented, descriptive statistics, along with appropriate graphical representations,⁵ are the first steps in the examination and analysis of data. Descriptive statistics topics in this chapter include the following:

- *measures of centrality, §2.3*
 - arithmetic mean*
 - median*
 - mode*
 - midrange*
 - trimmed mean*
 - Winsorized mean*
 - geometric mean*
 - harmonic mean*
 - weighted mean, §2.4*

⁵ Chapter 3 discusses graphical representations.

- *measures of dispersion, §2.6*
 range
 percentiles, quartiles, and quantiles
 interquartile range
- *variance, standard deviation, and degrees of freedom, §2.7*
- *statistics for coded data, §2.10*
- *measures of symmetry and peakedness, §2.11*
- *estimating portions of a dataset that lie in selected intervals, §2.12*
- *estimating the standard deviation of a dataset from its range, §2.13*
- *Chebyshev's inequality, §2.14*
- *spreadsheet functions for descriptive statistics, §2.5, §2.8*

2.2 Descriptive statistics

descriptive statistics Numeric values calculated from a dataset of values or one or more variables with the purpose of characterizing the behavior of the variable(s) are called ***descriptive statistics***. If the dataset is a random sample, the descriptive statistics often are the basis for making inferences about the population from which the sample is believed to have been drawn.

data reduction Note that sometimes the term ***data reduction*** is used in reference to descriptive statistics. Data reduction is the process of condensing a dataset into a few useful and informative statistics.

Descriptive statistics are present in many forms in daily life. To name a few:

- average wind velocity or the year's extreme temperatures at a nuclear plant
- minimum water temperature at a nearby lake
- maximum weight of fuel hauled by a tractor trailer
- proportion of funds allocated to research out of the total budget
- average cost of inspection (routine or ad hoc) of a facility
- average security cost for a facility or a group of facilities
- average and range of income of a starting nuclear engineer

Many commonly used descriptive statistics are easily calculated using a hand-held calculator or a computer spreadsheet program. Because descriptive statistics are so readily available and because the casual, even thoughtless, use of them may have serious consequences, it is important to make sure that only appropriate statistics are used.

2.3 Measures of centrality

measure of centrality A **measure of centrality** (*central value*, *central location*) is a quantity that is often used to describe a dataset, be it a population or a sample, by a single value. An example of such a measure is the average number of fuel assemblies that were shipped annually from a facility in the last 8 years. An often-used alternate term, **central tendency**, is inappropriate because the name suggests a measure of the proclivity of a set of values to cluster around a central value rather than a measure of the central value itself.

central value

central location

central tendency

This chapter presents several measures of centrality; choose among these carefully. Often, the mean (arithmetic mean) is the most intuitive and most useful measure of centrality. However, other measures may be preferred, depending on the type of data and the use to which the measure is put.

The following measures of centrality apply to any group of n values, regardless of whether that group is a sample or a population.

arithmetic mean The **arithmetic mean** of n values is the sum of the values divided by n .

mean Note that when there is no ambiguity, the arithmetic mean may be referred to as the **mean** or the **average**.

average

For the five values {83.44, 83.49, 84.11, 85.06, 85.31}, the mean is calculated as:

$$(83.44 + 83.49 + 84.11 + 85.06 + 85.31) / 5 = 421.41 / 5 = 84.282$$

(note the number of decimal places)

Consistent with accepted statistical practice, the mean is reported in the final presentation to one more decimal place than the data show. Intermediate steps, however, should be carried with as many significant figures as the computing facility permits.

Mathematically, the mean of n values, Y_1, Y_2, \dots, Y_n is denoted by \bar{Y} (pronounced “Y-bar”) and is defined as:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (2.1)$$

In Equation (2.1), Σ is the standard summation notation, and i is the index variable, running from 1 to n .

When there is no ambiguity, any of the following, less detailed, expressions in Equation (2.2) could be used:

$$\bar{Y} = \frac{1}{n} \sum_1^n Y_i = \frac{1}{n} \sum_i Y_i = \frac{1}{n} \sum Y = \frac{1}{n} \sum Y \quad (2.2)$$

- Note the convention: Variables are symbolized by upper case Latin letters such as Y , Y_i , or \bar{Y} . Once these values are known or observed, they are denoted by lower case letters such as $y_7 = 30.6$ or $\bar{y} = 0.06$.

median The **median** of a group of n values is the “middle” of that group. The definition of the median depends on whether n is odd or even.

The median of an odd number of ordered (listed from the smallest to the largest) values is defined as the middle value. As an example, for the dataset of $n = 5$ values {83.44, 83.49, 84.11, 85.06, 85.31}, the median (underscored here) is 84.11.

The median of an even number of ordered values is the average of the two middle values. As an example, consider the ordered set of $n = 6$ values {1.4, 1.7, 2.8, 3.2, 3.2, 4.6} (two middle values underscored). For this set, the median is calculated as $(2.8 + 3.2) / 2 = 3.0$. Note that, in this example, the number of decimal places for the median is the same as that for the raw data. However, if the middle values were, for example, 3.1 and 3.2, the median $(3.1 + 3.2) / 2 = 3.15$ would have one additional significant figure.

The median has several noteworthy properties. When doing hand calculations on ordered data, the median is easier to calculate (especially for a small value or values) than the mean because the middle value (or values) is often easy to spot. The median is also insensitive (“robust”) to extremely high or extremely low values. As an example, in the set {1.4, 1.7, 2.8, 3.2, 3.2, 4.6}, the median would still be 3.0, even if the largest value were a figure such as 460 rather than 4.6. For example, economic data are often summarized by medians (e.g., median salaries or home sale prices), because averages can be distorted by a few very high values.

The median is used in the construction of box plots (Section 3.9) and stem-and-leaf displays (Section 3.11).

mode The **mode** of a set of values is the measurement that occurs most often in the set. For the set of six values {1.7, 3.2, 3.2, 4.6, 1.4, 2.8}, the mode is 3.2 because 3.2 appears more often than any other value in the set. There is no mode for the set {83.44, 83.49, 84.11, 85.06, 85.31} because no value is repeated.

A dataset may have more than one mode. Thus, the set

Bimodal {2, 3, 3, 4, 4, 4, 4, 5, 6, 6, 7, 8, 8, 8, 8, 9, 11} has two modes:
trimodal 4 (appearing four times) and 8 (appearing four times). In this
multimodal example, we have a **bimodal** dataset. Similarly, a dataset may be
trimodal or **multimodal**.

Multimodality is not a useful measure of centrality. In fact, multimodality suggests that a single measure of centrality may be inappropriate because the sample could have been drawn from a mixture of several populations.

midrange The **midrange** of a set of values is the average of the smallest and the largest values in the set. For the set of values {1.4, 1.7, 2.8, 3.2, 3.2, 4.6}, where the smallest value (denoted here as y_{\min} and sometime as “minimum” or “min”) is 1.4 and the largest (y_{\max} , “maximum,” or “max”) is 4.6, the midrange is calculated as $(1.4 + 4.6) / 2 = 3.0$. Whereas the midrange may be easily calculated, it is sensitive to extreme values that are very large or very small.

trimmed mean The **trimmed mean** of a set of values is the average of the set after the smallest and the largest values have been removed. For the set of values {1.4, 1.7, 2.8, 3.2, 3.2, 4.6}, removal of the smallest (1.4) and the largest (4.6) values gives the trimmed mean as:

$$(1.7 + 3.2 + 3.2 + 2.8) / 4 = 10.9 / 4 = 2.725, \text{ rounded to } 2.72$$

rounding rule Note the **rounding rule**: Most statistics should be calculated to one more significant digit than the raw data. When the last digit before rounding is 5, round up or down to the nearest even integer. For example, 2.725 is rounded down to 2.72, whereas 2.735 is rounded up to 2.74.

An example of using a trimmed dataset is found in scoring for some competitions, where the highest and the lowest scores are removed in an attempt to eliminate or reduce judges’ biases.

Trimmed means are sometimes calculated after a certain proportion of the dataset’s smaller and larger values are removed. For example, a 10% trimmed mean is calculated after removing the smallest 5% and the largest 5% of values in the dataset.

Winsorized mean The **Winsorized mean** is a modification of the trimmed mean. If we replace the largest value in the dataset with the next-to-largest value and, symmetrically, the smallest value with the next-to-smallest value, the mean of the modified dataset is called the **first-level Winsorized mean**. (The adjective first-level refers to the replacement of only one value at each extreme of the data.) For

the set of values {1.4, 1.7, 2.8, 3.2, 3.2, 4.6}, the first-level Winsorized mean is calculated as:

$$(1.7 + 1.7 + 2.8 + 3.2 + 3.2 + 3.2) / 6 = 2.63$$

Both the trimmed and the Winsorized means are attempts to define a measure of centrality that is not overly influenced by extremes at either end of the dataset. Even though these extremes may be legitimate data points, they have no influence whatsoever on the trimmed mean. The Winsorized mean may allow the extremes to have some, although less, influence on the measure of centrality.

For a more complete discussion of trimmed and Winsorized means, see Dixon and Massey (1983), p. 380.

geometric mean The **geometric mean**, denoted here by G , is the n^{th} root of the product of n positive values. Mathematically,

$$G = \left(\prod_{i=1}^n Y_i \right)^{1/n} = \left(\prod Y_i \right)^{1/n} = \sqrt[n]{\prod Y_i} \quad (2.3)$$

Using logarithms (to any base!), Equation (2.3) can be written as:

$$\log G = \frac{1}{n} \sum \log Y_i \quad (2.4)$$

and G is calculated as the antilog of ($\log G$).

As an example of the geometric mean, suppose that three experts estimate the probability of a specific accident as 0.01, 0.001, and 0.00001. The arithmetic mean of these values is calculated as $(0.01 + 0.001 + 0.00001) / 3 = 0.00367$. The geometric mean is calculated as $\sqrt[3]{(0.01)(0.001)(0.00001)} = 0.000464$. We see that the geometric mean is considerably smaller than the arithmetic mean. In examples like this, the geometric mean is often considered a more appropriate measure of centrality because the arithmetic mean is much more influenced by the largest value in the dataset than is the geometric mean when dealing with numbers spanning several orders of magnitude.

harmonic mean Other measures of centrality are found in the literature, some of them developed for specific situations. One such measure is the **harmonic mean**, which is denoted here by H and is given by:

$$H = \frac{n}{1/Y_1 + 1/Y_2 + \dots + 1/Y_n} \quad (2.5)$$

Thus, the harmonic mean of the six values 1.4, 1.7, 2.8, 3.2, 3.2, and 4.6 is:

$$H = \frac{6}{1/1.4 + 1/1.7 + 1/2.8 + 1/3.2 + 1/3.2 + 1/4.6} = 6/2.502 = 2.40$$

Although the harmonic mean does not have many direct and practical applications, it has indirect uses. For example, it is an intermediate step in Duncan's Multiple Range Test (Section 16.8).

The definitions of measures of centrality discussed above are applicable to values coming from either a sample or a population. When the group values comprise the entire population, the arithmetic mean is called the **population mean** and is denoted by μ .

When the population contains a finite number of items, the size of the population is denoted by N , and the population mean is defined by any of the following equivalent expressions:

$$\mu = \frac{1}{N} \sum_{i=1}^N Y_i = \frac{1}{N} \sum Y_i = \frac{1}{N} \sum Y \quad (2.6)$$

That is, the population mean is the arithmetic average of the values of the N elements in the population. To calculate the population mean, add all the values in the population and divide the sum by N .

2.4 The weighted mean

The **weighted mean** arises in circumstances in which certain values in a dataset are considered to be more "important" than others and, therefore, are given more "weight" than others when the data are summarized. The weighted mean is a statistic constructed to accommodate these different weights.

Suppose we have k distinct Y_i values, each associated with its own weight w_i . The weighted mean is denoted as \bar{Y}_w , and calculated as:

$$\bar{Y}_w = \frac{\sum_{i=1}^k w_i Y_i}{\sum_{i=1}^k w_i} \quad (2.7)$$

Example 2.1 illustrates the use of the weighted mean in an academic institution, where some examinations carry more weight than other, less important examinations or quizzes.

Example 2.1. Final score in a course. The weights assigned to three classroom examinations are 20 percent to the first, 20 percent to the second, and

60 percent to the final. A student scored 70, 60, and 95 on his first, second, and final examination, respectively. Calculate the course score.

Data: 70, 60, 95; $k = 3$

Values: $y_1 = 70, y_2 = 60, y_3 = 95$

Weights: $w_1 = 20, w_2 = 20, w_3 = 60$

$$\sum_{i=1}^3 w_i = 100, \quad \sum_{i=1}^3 w_i y_i = (20)(70) + (20)(60) + (60)(95) = 8300$$

$$\bar{y}_w = 8300 / 100 = 83.00$$

Note that this example uses weights of 20, 20, and 60. However, the same result is obtained when one uses, respectively, weights of 2, 2, and 6, or of 1, 1, and 3. Thus, the ratios of the set of weights, rather than their specific values, are all that matter.

The weighted mean is also used when values are reported in intervals and each interval is represented by a single value \bar{y}_i , called the **class marker**, which may be the midpoint of interval i or another convenient value. The interval i is the i^{th} **class interval**. The number of values falling into interval i becomes the weight associated with y_i . For example, time to accident may be 71, 76, 78, and 78 days, but all four values (interval weight = 4) in the interval between 70 and 79 are reported as 74.5, 74, or 75, depending on how the class marker is determined.

Sometimes, a dataset is reported only as grouped data. Example 2.2 illustrates the calculation of the weighted mean for grouped data.

Example 2.2. Force to break. Eleven measurements of the force, in kilopounds (kip), required to break a specific weld, are listed and grouped into intervals. The unweighted mean of the $n = 11$ values is $\bar{y} = 5.42$.

Data: 2.1, 2.6, 2.7, 3.4, 5.2, 5.5, 5.9, 6.6, 7.4, 7.8, 10.4; $n = 11$; $k = 9$ intervals

The values are grouped into $k = 9$ intervals, each one kip wide, starting at 2 and ending at 11. The value for each interval is set at the midinterval as shown below.

Values: $y_1 = 2.5, y_2 = 3.5, y_3 = 4.5, y_4 = 5.5, y_5 = 6.5, y_6 = 7.5, y_7 = 8.5,$
 $y_8 = 9.5, y_9 = 10.5$

Weights: $w_1 = 3, w_2 = 1, w_3 = 0, w_4 = 3, w_5 = 1, w_6 = 2, w_7 = 0, w_8 = 0,$
 $w_9 = 1$

$$\sum_{i=1}^9 w_i = 11, \quad \sum_{i=1}^9 w_i y_i = 59.5, \quad \bar{y}_w = 59.5 / 11 = 5.41$$

As expected, the weighted mean based on interval grouping is a close approximation to the mean of the raw data. The difference between the unweighted and the weighted means results from the loss of some resolution in the process of grouping.

Still another use of the weighted mean is to expedite the calculation of the mean when some data are repeated. In this case, each distinct value y_i has an associated weight w_i , which is the number of repetitions of the weight of that value.

Example 2.3 illustrates the application of the weighted mean in calculating the average time before a metal container begins to rust.

Example 2.3. Time to rust. Data for the time (years) to first indication of rust taken on nine metal containers are shown below. Of the nine values, $k = 6$ values are distinct. The (unweighted) mean of the values is $\bar{y} = 5.30$.

Data: 2.3, 3.4, 3.4, 3.4, 5.2, 5.2, 6.6, 7.8, 10.4; $n = 9$

Values: $y_1 = 2.3, y_2 = 3.4, y_3 = 5.2, y_4 = 6.6, y_5 = 7.8, y_6 = 10.4; k = 6$

Weights: $w_1 = 1, w_2 = 3, w_3 = 2, w_4 = 1, w_5 = 1, w_6 = 1$

$$\sum_{i=1}^6 w_i = 9, \quad \sum_{i=1}^6 w_i y_i = 47.7, \quad \bar{y}_w = 47.7 / 9 = 5.30$$

Notice that the weighted mean is identical to the unweighted mean of all nine values. In this case, the weighted mean is simply another way to calculate the mean.

2.5 Spreadsheet functions for measuring centrality

Spreadsheets have special functions to calculate selected measures of centrality. Suppose that a given dataset is located in cells A1 through A100 of the spreadsheet. We illustrate the use of these functions in Excel, and call your attention to possible departures from other spreadsheet functions.

To calculate the mean of the dataset, use the function =AVERAGE(A1..A100).

➤ In some spreadsheets, AVERAGE is abbreviated as AVG.

To calculate the median of the dataset, use the function =MEDIAN(A1..A100).

To calculate the mode of the dataset, use the function =MODE(A1..A100).

- Warning: If more than one mode is found in the dataset, the spreadsheet will display only one of them.

To calculate the geometric mean of the dataset, use the function =GEOMEAN(A1..A100).

To calculate the harmonic mean of the dataset, use the function =HARMEAN(A1..A100).

No spreadsheet function yields the midrange of a dataset. However, the midrange can be easily calculated by summing the largest (y_{\max}) and the smallest (y_{\min}) values in the dataset and dividing the sum by 2. Specifically, $=(\text{MAX}(A1..A100)+\text{MIN}(A1..A100))/2$.

Excel does not give the measures of centrality for trimmed data. To obtain a trimmed mean or a Winsorized mean, perform the trimming by hand first, and then apply the average to the trimmed dataset. Similarly, there is no single command function to calculate the weighted mean. The weighted mean may be calculated by entering the values and the weights in two columns, calculating and adding their cross products, and dividing the latter by the sum of the weights.

2.6 Measures of dispersion: range and quantiles

measure of dispersion
spread
variability

A measure of centrality—be it mean, median, mode, or some other—is generally considered to be the most important parameter/statistic associated with a set of values. However, a measure of centrality does not tell the whole story about a variable. For a more complete description of a variable, some *measure of dispersion (spread, variability)* of the data is required.

range
quantiles
percentiles
quartiles

Measures of dispersion described in this section are the *range*, and functions of *quantiles* are points taken at regular intervals from the cumulative distribution function (such as *percentiles* and *quartiles*). Section 2.7 discusses the variance and the standard deviation, which are also measures of dispersion.

The range is defined as the difference between the largest value (Y_{\max}) and the smallest value (Y_{\min}) in a set of values. Formally,

$$\text{range} = Y_{\max} - Y_{\min} \quad (2.8)$$

As an example, in the ordered dataset $\{1.4, 1.7, 2.8, 3.2, 3.2, 4.6\}$, the largest value (y_{\max}) is 4.6 and the smallest value (y_{\min}) is 1.4, so that the range is $y_{\max} - y_{\min} = 4.6 - 1.4 = 3.2$.

The range is an intuitive measure of data dispersion, as a large (or small) range suggests large (or small) variability. If the largest and the smallest values are easy to spot, the calculation of the range is very easy.

We may have a sample range, and, in parallel, there may be a population range. However, for some populations and samples, the range cannot be determined. This may be caused by the lower detection limit, the higher detection limit, or other instrument or technician limitations.

- The sample range cannot be larger than the population range; indeed, very seldom are the two equal.

Percentiles apply to any set of values on an ordinal or an interval scale. As the term suggests, they are based on the division of an ordered (from smallest to largest) dataset into 100 equal parts. However, if the number of values in the dataset is not a multiple of 100, there may be no clear-cut definition of percentile. One definition in use is the following:

For any p , $0 \leq p \leq 1$, the ($100p^{\text{th}}$) percentile is a value (not necessarily in the dataset) such that at most ($100p$) percent of the values are less than this value and at most $100(1 - p)$ percent are greater.

One problem with this definition is that a percentile may not have a unique value. For example, consider the set $\{1.0, 2.0, 3.0, 4.0\}$, where any value between 2.0 and 3.0 can qualify as a 50th percentile. Another problem is that a single value may represent a range of percentiles. For example, what is the 80th percentile of this dataset? From the second part of the definition, at most 20% of the values are greater. Therefore, the 80th percentile is 4.0, the largest value. Similarly, the 76th to the 100th percentiles are all equal to 4.0.

To avoid these problems with the definition of percentile, it is necessary to interpolate between the values in the dataset. However, as Example 2.4 illustrates below in a slightly different context, there is more than one way to interpolate. All the interpolation methods yield similar results, and no method appears to have a distinct advantage over others.

A quartile is one of three particular percentiles of special interest in describing sets of data:

- The first quartile is the 25th percentile of a dataset.
- The second quartile is the 50th percentile of a dataset. The second quartile is also the median. (See Section 2.3. The median as defined in

Section 2.3 will always be equal to the 50th percentile as defined below.)

- The third quartile is the 75th percentile of a dataset.

interquartile range A common measure of dispersion is the **interquartile range**, defined as the difference between the first and third quartiles. We may think of it as the range of the middle half of the data. The interquartile range is used in the construction of box plots (Section 3.9).

quintiles Clearly, the quartiles divide the dataset into four approximately equal parts. Division points related to the quartiles are the **quintiles** (five approximately equal parts) and the **deciles** (10 approximately equal parts).

deciles

Quantile is the collective term for the percentiles, deciles, quartiles, and quintiles, or other “-iles.” A quantile is often written as a decimal fraction. Thus, we define the 0.85 quantile of a set of data to be a value on the scale of the data that divides the data into two groups, so that a fraction of about 0.85 of the values falls at or below that value and a fraction of about 0.15 falls above it. Such a value is often denoted as $Q(0.85)$. The only difference between percentile and quantile is that percentile refers to a percentage of the dataset, and quantile refers to a fraction of the dataset.

- A quantile multiplied by 100 yields the corresponding percentile.
- A percentile divided by 100 yields the corresponding quantile.

fractile Some writers use the word **fractile** instead of quantile, perhaps to emphasize that it is a quantity (fraction) between 0 and 1.

Example 2.4. Water impurity. Twenty-five aliquots of water were taken at random locations from an aquifer to establish a baseline for water impurity in parts per million (ppm). Table 2.1 lists the 25 values in ascending order. Suppose, for regulatory purposes, we need to calculate $Q(0.85)$, the 85th quantile of the dataset. Since none of the values serves to set a value for $Q(0.85)$, some arbitrariness cannot be avoided, as different approaches lead to different results. Since 0.85 of 25 is 21.25, $Q(0.85)$ should be between 42.91 and 42.93, the 21st and 22nd ordered values, respectively. The fractional part of 21.25 suggests that the 85th quantile be 0.25 of the distance between the 21st and 22nd ordered values. Thus, $Q(0.85)$ is calculated as $42.85 + (0.25)(42.91 - 42.85) = 42.865$. Alternatively, $Q(0.85)$ can be calculated as $(0.75)(42.85) + (0.25)(42.91) = 42.865$. The equivalence of these two alternative calculations can be verified algebraically.

Table 2.1. Water impurities from 25 aliquots, (PPM)

42.31	42.32	42.40	42.42	42.45
42.45	42.48	42.49	42.50	42.59
42.61	42.61	42.62	42.63	42.65
42.67	42.72	42.82	42.83	42.83
42.85	42.91	42.93	42.97	43.17

A second method to calculate quantiles of a dataset is the following. For a dataset with n values, rearrange the values in increasing order of magnitude. Denote the ordered values with subscripts in parentheses so that $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$. The p^{th} quantile is then defined as follows:

For any p , $0 \leq p \leq 1$, set $p(n+1) = k + d$, where k is an integer and $0 \leq d \leq 1$. Denote the $(100p^{\text{th}})$ percentile by $Y(p)$. Then,

$$Y(p) = Y_{(k)} + d(Y_{(k+1)} - Y_{(k)}) \text{ for } 0 < k < 1$$

$$Y(p) = Y_{(1)} \text{ for } k = 0$$

$$Y(p) = Y_{(n)} \text{ for } k = n$$

With this definition, every percentile has a unique value, although two or more percentiles can have the same value if p is close to 0 or 1, or if two or more consecutive values in the dataset are equal.

In Example 2.4, where $n = 25$, we see that 0.85 of $n + 1 = 26$ is 22.10, and $Q(0.85)$ should be 10 percent of the way between the 22nd and 23rd ordered values. Thus, $Q(0.85) = 42.91 + (0.10)(42.93 - 42.91) = (0.90)(42.91) + (0.10)(42.93) = 42.912$. This shows that the second method of calculating quantiles or percentiles can yield different values than those obtained by the first method.

Finally, using Excel's =PERCENTILE(data range, 0.85) function, we get $Q(0.85) = 42.874$. This value is the result of a compromise between the two methods discussed above.

- In practice, any of these methods for calculating percentiles is acceptable.
- In computer spreadsheets, all “-iles” must be entered as a quantity between 0 and 1. Thus, the 67th percentile is entered as 0.67, the third quartile as 0.75, the second quintile as 0.40, and the seventh decile as 0.70. See Section 2.8 for the explicit function to be used in a spreadsheet.

2.7 Variance and standard deviation

variance The **variance** and the **standard deviation** are the most common measures of dispersion. These measures are defined first for populations and then for samples.

variance of a finite population The **variance of a finite population** of size N , with mean μ and values $\{Y_i, i = 1, \dots, N\}$, is defined by:

$$\sigma^2 = \frac{\sum_{i=1}^N (Y_i - \mu)^2}{N} \quad (2.9)$$

population standard deviation The **standard deviation of a population** is the positive square root of the variance. Mathematically,

$$\sigma = \sqrt{\sigma^2} \quad (2.10)$$

- The standard deviation is generally easier to relate to than the variance because the standard deviation is expressed in the same units of measurement as the original data, whereas the variance is expressed in the square of those units.

The variance and standard deviation of a sample are measures that imitate their population counterparts.

sample variance For a sample of size n , the **sample variance** is defined by one of two expressions, according to whether we know (either by historical data, by manufacturer specifications, or by assumption) or do not know the value of the mean μ of the population from which the sample is drawn.

On those rare occasions when μ is known, the sample variance is:

$$s^2 = \frac{\sum (Y_i - \mu)^2}{n} \quad (2.11)$$

For example, if $n = 6$ sample values are 1.4, 1.7, 2.8, 3.2, 3.2, 4.6, and the population mean is known to be 2.5, the sample variance is calculated by Equation (2.9) as:

$$\begin{aligned} s^2 &= [(1.4-2.5)^2 + (1.7-2.5)^2 + (2.8-2.5)^2 + (3.2-2.5)^2 + (3.2-2.5)^2 + (4.6-2.5)^2] / 6 \\ &= 7.330/6 = 1.22. \end{aligned}$$

When μ is not known, μ is estimated by the sample mean \bar{Y} , and the sample variance is calculated as:

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} \quad (2.12)$$

Equation (2.12) is called the definition formula for the sample variance.

An alternative to Equation (2.12), called the working formula for the sample variance, can be written in several forms, all of which are mathematically equivalent to Equation (2.12):

$$S^2 = \frac{\sum Y_i^2 - (\sum Y_i)^2 / n}{n-1} = \frac{\sum Y_i^2 - n\bar{Y}^2}{n-1} = \frac{n\sum Y_i^2 - (\sum Y_i)^2}{n(n-1)} \quad (2.13)$$

Similar to the population standard deviation, the standard deviation for the sample is the positive square root of the sample variance:

$$S = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n-1}} \quad (2.14)$$

The denominator of S^2 , either n when μ is known (Equation (2.11)) or $n - 1$ when μ is not known (Equation (2.12)), is called the **degrees of freedom**. Degrees of freedom is often denoted by the Greek letter nu (ν).

- Degrees of freedom is generally used as a singular noun. It corresponds to the number of independent observations on which the sample statistic is based. Degrees of freedom is seldom equal to the number of data points.
- It can be puzzling that the denominator of the sample variance is divided by $n - 1$ (degrees of freedom) rather than n . An explanation is that, while the numerator of S^2 is the sum of n deviations squared, the deviations sum to zero. Thus, if $n - 1$ of the deviations were freely chosen, the n^{th} one is obtained by subtraction, so that S^2 is based on only $n-1$ independent observations.

2.8 Spreadsheet functions for measuring dispersion

Spreadsheets have special functions to calculate selected measures of dispersion. Again, without loss of generality, assume that a given dataset is located in cells A1 through A100 of the spreadsheet. We illustrate the use of these functions in

Excel and call your attention to possible differences from other spreadsheet notations.

As stated earlier, there is no direct spreadsheet function to calculate the range, but it can be calculated by subtracting the smallest value from the largest value in the dataset:

`=MAX(A1..A100) – MIN (A1..A100)`

Also, as discussed earlier, the quantile is calculated using the percentile function. So, for the 85th quantile, we use the function:

`= PERCENTILE(A1..A100, 0.85)`

- Note that the quantity q must be entered as a decimal fraction between 0 and 1.

When using spreadsheet software, be aware that some built-in functions may be confusing in their notations. Note, for example, that the VAR function is not the same for Excel and some other spreadsheet programs, so stay alert!

To find the variance of a population (or of a sample when μ is known), use the function `=VARP(A1..A100)`.

- In some spreadsheets, the population variance is written as VAR.

To find the variance of a sample (μ unknown), use the function `=VAR(A1..A100)`.

- In some spreadsheets, the sample variance is written as VARS.

To find the standard deviation of a population, use the function `=STDEVP(A1..A100)`.

- In some spreadsheets, the population standard deviation is written as STD.

To find the standard deviation of a sample, use the function `=STDEV(A1..A100)`.

- In some spreadsheets, the sample standard deviation is written as STDS.

2.9 Descriptive statistics with a hand-held calculator

A personal (hand-held) calculator may be all we need to generate the necessary descriptive statistics, especially when the dataset is small. Unfortunately, the

notation, or symbols, displayed on the calculator may differ from one calculator to another.

The average may be labeled as \bar{x} , \bar{y} , or “avg.” But don’t take our word for it; check the manual.

The variance and standard deviation may also be calculated on a scientific calculator. Just be careful that you know which button to push, as it may be confusing. Some personal calculators use σ_n and σ_{n-1} to denote the standard deviations for the population and the sample, respectively. Some calculators use σ and s to make the same distinction. Other calculators fail to differentiate between the two in any way. To see what a given calculator does, enter the values 1, 2, and 3 and press the standard deviation key. If the display shows exactly 1.000, the calculator uses $(n - 1)$ as the denominator. If the display shows a value in the neighborhood of 0.816, the calculator uses n as the denominator. If any value other than 1.000 or 0.816 appears in the display, check the operator’s manual or buy another calculator (no, it’s not that the batteries have gone bad!).

Note that personal calculators use the working formula for the variance calculations—any of the equations of the Equation (2.13) variety. On rare occasions, the working formula may lead to serious errors, depending on the number of digits (“working digits”) the calculator uses in its operation. If the entered data have many significant digits and those entries are squared and summed, the calculator may have an “overflow” that will give a completely erroneous result.

For example, in containment integrated leakage-rate testing, the air mass, measured in pounds, may be written with 10 significant digits. When such values are squared and added, the results are unpredictable. One of the solutions is to code the data, as discussed in the next section.

2.10 Descriptive statistics for coded data

Data coding Spreadsheets, like hand-held calculators, are not entirely immune to the problem of limited precision. Spreadsheets, however, use more significant digits and are less likely to encounter calculation errors. **Data coding** is one way to overcome the problem of insufficient working digits. Consider a set of n values $\{y_1, y_2, \dots, y_n\}$ with a mean \bar{y} and a standard deviation S . Multiply every y_i by a constant c_1 and then add another constant c_2 .

The resultant coded values, $\{x_1, x_2, \dots, x_n\}$, can be written as:

$$x_i = c_1 y_i + c_2, \quad i = 1, \dots, n \quad (2.15)$$

The mean and standard deviation of the x_i are $\bar{x} = c_1 \bar{y} + c_2$ and $S_x = c_1 S$, respectively. Note that the addition operation does not affect the standard deviation. Once the mean and standard deviation of the coded data $\{x_i\}$ are calculated, a reverse transformation is made to obtain values for \bar{y} and S . The mean and standard deviation of $\{y_i\}$ are equal to $(\bar{x} - c_2) / c_1$ and S_x / c_1 , respectively.

Example 2.5. Containment air mass. On one day, 33 measurements of air mass are collected at random from inside containment for the purpose of establishing a baseline. The mean and standard deviation of the raw data are calculated as 173808.52 and 11.82, and counterpart statistics from the coded data are 8085.2 and 118.2. From the coded statistics, the raw data statistics are restored as $\bar{x} = 8085.2 / 10 + 173000 = 173808.52$. The standard deviation of the raw data is $118.2 / 10 = 11.82$. The last value should be squared to get the variance.

Table 2.2 presents the raw data and the coded data, where each containment air mass value is first transformed by adding a negative constant $c_2 = -173000$ and then multiplying by a constant $c_1 = 10$.

The mean and standard deviation of the raw data are calculated as 173808.52 and 11.82, and counterpart statistics from the coded data are 8085.2 and 118.2. From the coded statistics, the raw data statistics are restored as $\bar{y} = 8085.2 / 10 + 173000 = 173808.52$. The standard deviation of the raw data is $118.2 / 10 = 11.82$. The last value should be squared to get the variance.

Table 2.2. Data coding containment air mass measurements

Point	Air mass (lb)	Coded data	Point	Air mass (lb)	Coded data	Point	Air mass (lb)	Coded data
1	173825.3	8253	12	173799.7	7997	23	173806.5	8065
2	173817.0	8170	13	173817.2	8172	24	173797.8	7978
3	173801.9	8019	14	173829.0	8290	25	173812.4	8124
4	173828.0	8280	15	173814.5	8145	26	173809.1	8091
5	173785.7	7857	16	173788.0	7880	27	173819.1	8191
6	173798.5	7985	17	173819.3	8193	28	173808.0	8080
7	173800.3	8003	18	173815.3	8153	29	173820.1	8201
8	173795.3	7953	19	173815.2	8152	30	173806.4	8064
9	173804.2	8042	20	173800.6	8006	31	173794.7	7947
10	173822.3	8223	21	173822.9	8229	32	173817.8	8178
11	173794.7	7947	22	173799.1	7991	33	173795.2	7952
Coded data: mean = 8085.2,			standard deviation = 118.2					
Raw data: mean = 173808.52,			standard deviation = 11.82					

2.11 Skewness and kurtosis

skewness **Skewness** is a measure of the lack of symmetry about the mean. When a distribution is symmetric, the coefficient of skewness is zero. Positive skewness is associated with a long “tail” or “drag” at the right of the distribution. Negative skewness is associated with a long tail at the left of the distribution. Larger numerical skewness values are associated with stronger departures from symmetry.

The sample coefficient of skewness is defined as:

$$m_3 = \frac{\sum (Y - \bar{Y})^3}{nS^3} \quad (2.16)$$

- Skewness is sometimes used as one of several criteria in testing for normality. However, better procedures for testing normality are available (see Chapter 11).
- Some writers consider a distribution to be essentially symmetric if $|m_3| < 0.5$.
- The definition of skewness is not unique. Other definitions may be found in Zwillinger and Kokoska (2000), p. 18.

kurtosis **Kurtosis** is a measure of the peakedness of a distribution. The sample coefficient of kurtosis is defined as:

$$m_4 = \frac{\sum (Y - \bar{Y})^4}{nS^4} \quad (2.17)$$

For a large sample size (say, $n > 20$), the following applies:

- For a sample from a normal distribution (see Chapter 6), m_4 is about 3.
- If $m_4 < 3$, the sample is flatter than is expected from a normal distribution.
- If $m_4 > 3$, the sample has more of a peak than is expected from a normal distribution.
- The definition of kurtosis is not unique. Some authors define kurtosis as $m_4 - 3$, where m_4 is defined by Equation (2.17). Still other definitions of kurtosis, along with worked examples, can be found in Zwillinger and Kokoska (2000), p. 18.

- Kurtosis is sometimes used jointly with skewness to test whether a distribution is normal. However, skewness and kurtosis are of limited use in testing for normality, as more efficient procedures for such testing are available.

2.12 An empirical rule for a mound-shaped dataset

Many sets of measurements yield mound-shaped graphs when displayed as a histogram.⁶ That is, when all the measurements are plotted as a histogram, where the horizontal axis represents the measurements and the vertical axis gives the frequencies (or the relative frequencies) of the measurements, the result is a relatively symmetric, relatively smooth curve with tapering tails to the left and to the right of its center.

Histograms of datasets are seldom truly symmetric with smoothly tapering tails. However, the histograms of many datasets are roughly mound shaped (see Ott and Mendenhall (1984), p. 76).

For a histogram that is mound-shaped:

The interval from $(\bar{Y} - S)$ to $(\bar{Y} + S)$ contains approximately 68% of the measurements. (2.18)

The interval from $(\bar{Y} - 2S)$ to $(\bar{Y} + 2S)$ contains approximately 95% of the measurements. (2.18a)

The interval from $(\bar{Y} - 3S)$ to $(\bar{Y} + 3S)$ contains all, or nearly all, of the measurements. (2.18b)

This empirical rule provides convenient estimates of portions of a dataset falling in selected intervals, as illustrated by the following example.

Example 2.6. Weights of uranium ingots. Suppose we have 150 uranium ingot weights, as displayed in Table 2.3 (taken from Bowen and Bennett (1988), pp. 12).

⁶ See Section 3.7 for a discussion of histograms.

Table 2.3. Weights of 150 uranium ingots, in kilograms

425.0	426.7	423.3	429.4	427.9	425.9	422.1
422.5	424.4	427.8	428.9	425.3	425.4	427.3
424.9	422.4	426.1	424.8	432.4	427.9	421.9
431.7	432.2	422.4	427.3	427.3	423.7	425.7
426.3	427.8	424.8	428.0	426.3	428.6	425.9
424.5	431.3	431.2	427.3	418.5	428.8	431.6
426.1	425.8	429.8	429.5	425.3	424.5	424.6
423.1	426.8	430.9	423.9	421.9	425.1	421.8
428.3	424.8	427.0	425.1	425.2	424.4	432.3
423.2	423.6	427.9	427.9	428.5	424.7	428.8
428.2	424.8	421.0	423.6	428.0	427.7	425.7
429.1	429.7	419.6	421.3	426.8	421.2	425.2
424.2	430.3	424.6	430.0	423.5	427.2	430.0
429.7	423.2	428.8	425.4	427.5	429.4	424.9
424.8	431.0	427.9	423.6	421.7	425.9	426.6
427.2	428.0	428.0	429.7	427.4	426.6	426.2
428.3	426.6	428.4	427.1	427.5	425.5	426.2
429.3	425.4	423.1	426.9	425.7	429.2	434.2
421.8	427.3	425.2	427.3	425.7	426.5	420.4
424.0	426.0	424.9	430.5	426.3	426.3	424.9
428.0	423.3	431.1	426.4	429.0	429.9	423.3
427.4	424.2	428.2				

Figure 2.1 shows the histogram of these weights, which is mound-shaped. The mean is $\bar{y} = 426.40$, the standard deviation is $s = 2.83$, and the range is 15.7.

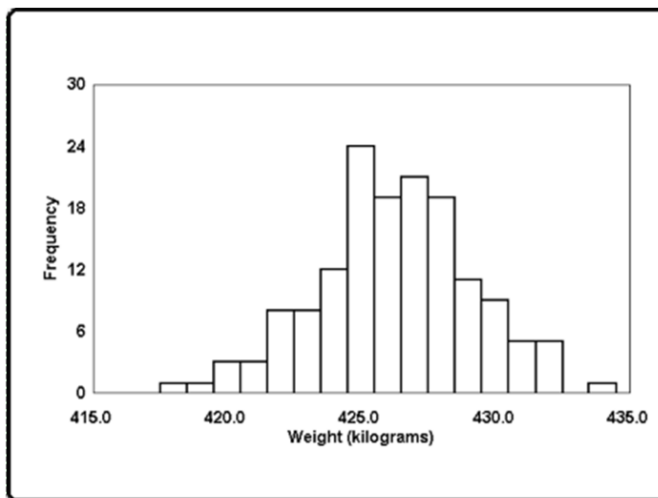


Figure 2.1. Histogram of uranium ingot weights in Table 2.3

Table 2.4 displays the grouping of the uranium ingot data into intervals and their corresponding percentages. Note the close agreement between the sample percentages and the empirical rule.

Table 2.4. Actual and approximate contents of intervals

Interval defined by	Evaluates as	Contains by the empirical rule (approximately)	Sample contains
$(\bar{Y} - S), (\bar{Y} + S)$	(423.58, 429.23)	68%	69.3%
$(\bar{Y} - 2S), (\bar{Y} + 2S)$	(420.75, 432.06)	95%	95.3%
$(\bar{Y} - 3S), (\bar{Y} + 3S)$	(417.93, 434.88)	100%	100%

2.13 Estimating the standard deviation from the range

Related to the mound-shaped conditions that define the empirical rule in Section 2.12 is an inequality involving the ratio of the range and the standard deviation that may hold for many sets of data:

$$3 \leq \frac{\text{range}}{\text{standard deviation}} < 6 \quad (2.19)$$

This relationship could be used to check on the calculation of the standard deviation. For example, if the ratio of the range to the standard deviation is 10, this suggests that something is either extremely unusual or just plain wrong.

- It can be shown mathematically that no matter what the shape of the distribution, the ratio of the range to the standard deviation cannot be smaller than 1.4142 (the square root of 2). So if the ratio of the range to standard deviation is very small, further investigation is warranted. In a pinch, one could use another empirical rule to obtain rough estimates of the standard deviation from the range, as shown in Table 2.5.

We applied this rule to the uranium ingot data in Table 2.3, where $n = 150$ and the range = 15.7. The rough estimate of the standard deviation is calculated as $15.7 / 6 = 2.62$. Recall that the correct standard deviation is 2.83.

These two empirical rules provide useful, albeit informal, computational and/or “eyeball” checks in many statistical investigations. However, because they do not always apply to every set of data (not even to those that are mound shaped), caution is needed in their use. They should never be used as substitutes for exact calculations in formal studies. Indeed, many writers frown on the use of this rule, and we share their sentiments for other-than-informal use.

Table 2.5. A rough estimate of the sigma from the range

Measurement set	Estimate
Small (say, 20 measurements or fewer)	Range / 4
Large (say, between 20 and 100 measurements)	Range / 5
Very large (say, more than 100 measurements)	Range / 6

2.14 Chebyshev's inequality

Chebyshev's inequality One of the best known theorems in probability is called *Chebyshev's inequality*, which is a rule of thumb to characterize spread in a dataset. It provides an upper bound for the probability that the values in a dataset or distribution differ from its mean in units of its standard deviation. While this inequality is mathematically correct, it is rarely of any practical use and is given here only for the sake of completeness.

Let Y be an arbitrary dataset or distribution with mean μ and standard deviation σ , and let $k > 0$ be a constant. Then,

$$Pr\{|Y - \mu| > k\sigma\} < 1/k^2 \quad (2.20)$$

Table 2.6 shows Chebyshev's probability bounds for selected values of k .

Table 2.6. Chebyshev's selected probability bounds

Multiple k of σ	Probability content
$k = 1$	$Pr\{ Y - \mu > \sigma\} < 1$ This statement, obviously, provides useless information. (2.21)
$k = 2$	$Pr\{ Y - \mu > 2\sigma\} < 1/4 = 0.25$ (2.22)
$k = 3$	$Pr\{ Y - \mu > 3\sigma\} < 1/9 = 0.11$ (2.23)

The approximate probability content from the empirical rule for mound-shaped distributions and the probability bound of Chebyshev's inequality are checked against the data from Table 2.3. Table 2.7 summarizes the results. The empirical rule provides an excellent approximation, while Chebyshev's inequality is of little use.

Table 2.7. Comparison of probability content

Interval defined by	Evaluates as	Content from the empirical rule for mound-shaped distributions	Content from Chebyshev's inequality	Actual sample content
$(\bar{y} - s), (\bar{y} + s)$	(423.58, 429.23)	68%	> 0%	69.3%
$(\bar{y} - 2s), (\bar{y} + 2s)$	(420.75, 432.06)	95%	> 75%	95.3%
$(\bar{y} - 3s), (\bar{y} + 3s)$	(417.93, 434.88)	100%	> 89%	100%

3

Statistical graphics

3.1 What to look for in Chapter 3

Graphical presentations are essential to almost any statistical analysis of data. Analysts often recommend that the data be plotted before the data analysis begins as the visual inspection may identify patterns and irregularities in the data and, consequently, may suggest an approach to the analysis.

This chapter presents some of the popular statistical graphs used to convey information, including the following:

- *pie chart*, §3.3, §3.4
- *bar chart*, §3.5, §3.6
- *line plot*, §3.5
- *histogram*, §3.7, §3.8

Two special-purpose data presentations are motivated and demonstrated:

- *box plot*, § 3.9, §3.10
- *stem-and-leaf display*, §3.11, § 3.12

Simplified guidelines for the construction of these statistical graphics are offered. Also, wherever Excel routines are available, we list Excel's relevant steps for such construction.

The discussion of graphical presentation is not restricted to this chapter. Other statistical graphics, such as scatter plots and regression plots, are presented in the contexts in which they are introduced. Furthermore, the graphical methods presented in this and other chapters do not begin to cover the many methods and improvisations thereof that are available in the literature, in software packages, and in the minds of professionals and laypersons.

3.2 What's in a graph?

Since the dawn of human recordkeeping on the walls of caves, graphical methods and statistical procedures have been intertwined. Charts of all types, using all sorts of colors and designs and even combining with sound effects, are part of the fabric of modern life. In a variety of forms, some certainly more useful than others, statistical graphics appear on television, on the Internet, and in newspapers, magazines, and other media. As simple as some of these chart types may be, they scarcely can be equaled as powerful devices to convey information contained in a collection of data.

Because statistical graphics are so powerful and so influential, they are subject to a range of ill use, from the deliberate hiding of facts to the inadvertent "chart junk." To use a concept emphasized by Tufte (1983), "They can tell the truth, the whole truth, and nothing but the truth—and yet mislead the reader." Statistical graphics can lead to under-interpretation (as when one fails to see a pattern) or to over-interpretation (where one perceives a pattern that is not real) of a study. Thus, honesty, integrity, caution, and introspection must guide the selection and building of graphs to convey a message.

Quality statistical graphics are more than just a means to present data. Essential statistical graphics must:

- display their source data in a readily apparent fashion
- be "self contained," as much as possible (i.e., convey the message without a need for additional text or supplementary material)
- contain all the necessary axis labels, axis units, and applicable captions
- be free of irrelevancies
- be simple and not overwhelm with details
- identify elements of the study (such as potential rounding errors) that may not be obvious to every reader

Whether deliberately or not, charts may be misleading. We need to examine a chart carefully as we examine data. Huff's criteria may be a good start for such examination.

A classic source on graphical presentation of quantitative information is a book by Tufte (1983) entitled *The Visual Display of Quantitative Information*. An interesting exposition on the vulnerability of the public to bad graphs may be found in *How to Display Data Badly* by Wainer (1984), p. 137.

The menu of potential graphical representations is wide and varied. The selection from this menu must be made with the recognition of the audience and the level of sophistication that the study warrants. Thus, we were cautioned before and we are cautioned again: Incidents where graphs distort facts (deliberately or not) abound. Such graphs may be self-serving. They may be derived from real data but represent an incomplete truth (as in the case of the Challenger disaster as discussed in Chapter 1).

Finally, we humbly suggest that you stick to traditional charts and avoid fancy or "creative" charts. Creative charts may distort reality. If the same information can be presented in two as well as three dimensions, choose the two-dimensional presentation. Similarly, an oval chart may look prettier than a plain, round pie chart, but it has the potential to mislead the reader.

Most statistical packages and spreadsheet programs have an extensive menu for selecting graphs and charts. Use keen judgment and good taste to select a display that is both appropriate and appealing.

3.3 The pie chart

pie chart A *pie chart* is a graphical display designed to show and sometimes emphasize the relative proportions of several values of a nominal scale where each value is assigned to one and only one group or category. This is illustrated by the following example

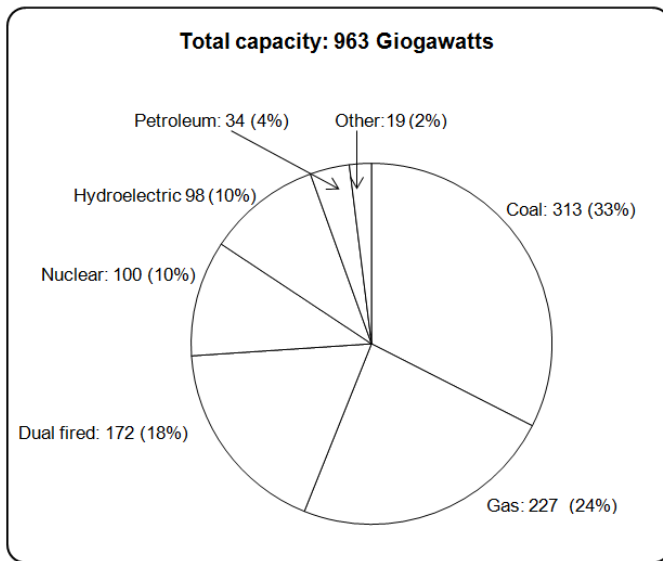
Example 3.1 Energy source distribution. Table 3.1 is a typical set of data that is often displayed by a pie chart. For each of several values of a variable on a nominal scale—in this case, the energy source (e.g., coal, gas)—there is an associated value of an interval scale—in this case, the amount of energy. We convert each of these values to a percentage (or proportion) of their total. Then we convert those percentages into proportional segments of a circle to form the "slices of the pie." Figure 3.1 presents the data in Table 3.1 as a pie chart.

Table 3.1. 2005 U.S. electric capacity by energy source

Source	Gigawatts	Percent of total
Coal	313	33
Gas	227	24
Dual fired*	172	18
Nuclear	100	10
Hydroelectric	98	10
Petroleum	34	4
Other	19	2
Total	963	100

* Dual-fired units can burn oil or gas.

Note: Due to rounding, the total may not add up to 100 percent.

**Figure 3.1. Pie chart derived from data in Table 3.1**

A pie chart, typically, does not stand by itself. The status quo may be interesting, but no less interesting is any change relative to other configurations, such as the electric capacity for another country, or to some previous year. Two or more pie charts may be drawn next to each other to show the changes.

Despite the pie chart's popularity, venerability, and ubiquity, it is not universally endorsed as a statistical display of data. For example, Stein, et al. (1985) lists the pie chart, among others, as not recommended. They go on to say, specifically:

The pie chart has the distinct advantage of indicating the data must add up to 100 percent. At the same time, it is difficult to see the relative size of two slices, especially if they are within 10 percent of each other.

And, in a similar vein, statisticians often cite the following maxim:

A pie chart is never better than the second best way to plot data.

With this caveat in mind, we continue with suggestions for the construction of pie charts.

3.4 Suggestions for constructing pie charts

The following rules are intuitive and consistent but sometimes should be broken:

1. The number of slices of the pie should be between 4 and 10. Having too many slices clutters the chart, while having too few slices may appear trivial and insult the reader.
2. If necessary, collapse several groups into a single category to reduce the number of slices. If necessary, combine them as “Other.”
3. Arrange the slices in increasing or decreasing order of magnitude, with the possible exception of “Other.” “Other” should be the very last category, unless there are compelling reasons to place it elsewhere.
4. Begin the first slice at the 12 o’clock position, followed clockwise by the other slices.
 - Counter-clockwise direction works as well; just be consistent.
5. Show the raw data as well as the percentages. Remember that if the raw data are given, percentages can be calculated and verified. In contrast, if only percentages are given, the data cannot be reconstructed.
 - Caution: Avoid using a pie chart for ordered categories because of the difficulty of seeing the ordering around a circle.

Pie charts may be constructed using Excel by following the steps listed below. Details of the many associated options may be found by clicking Excel’s help icon and panel.

1. List the category labels in one column. Thus, in Example 3.1, we may list the seven energy sources (“Coal”...“Other”) in ascending or descending order in cells A1..A7.

2. List the corresponding numerical values next to their labels in the next column—in this example, in cells B1..B7.
3. Select the cells that include the labels and the values. In this example, select cells A1..B7.
4. Click the graph icon. If the graph icon is not visible, click **Chart** in the **Insert** menu.
5. In the **Chart type** box, click **Pie**.
6. Click **Finish** to view the pie chart.
7. Embellish the chart by selecting frame, color, font, shading, and other options.
8. Be consistent, especially if you display more than one pie chart.

3.5 The bar chart

bar chart A *bar chart* is a graphical representation used to display nominal or ordinal data. Example 3.2 illustrates the use of the bar chart.

Example 3.2. Monthly rejects of fuel rods. The monthly number of fuel rods rejected in a fuel facility is listed for a full year (Bowen and Bennett, 1988, p. 12). This example illustrates features of the bar chart relative to the data displayed in Table 3.2.

Table 3.2. Monthly rejects of fuel rods

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rejects	5	4	4	2	4	7	6	2	2	5	4	4

Table 3.2 contains a set of data that typically leads to a bar chart, such as that shown in Figure 3.2. Each month has a frequency of occurrences. Each bar in the bar chart is intended to represent 1 month, with the bar's height representing the month's frequency of rejects. By default, the width of the bars is constant, as is the space between each pair of consecutive bars. If either of these two defaults is violated, the bar chart's message is subject to misinterpretation. If the categories are on an ordinal scale (as they are in Table 3.2), the bars are placed in sequence, usually chronologically or from the smallest to the largest. The bar chart may be oriented horizontally (sometimes called a *row chart*) or vertically (sometimes called a *column chart*). Figure 3.2 renders the data from Table 3.2 as a vertical bar chart.

row chart

column chart

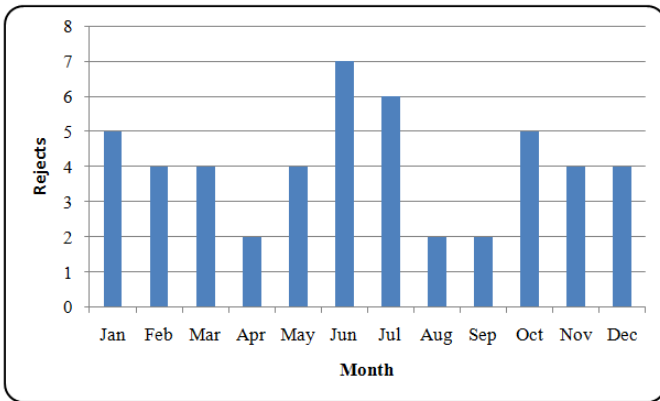


Figure 3.2 Bar chart for monthly fuel rod rejects

line plot The same data can be presented in a different display called a *line plot*, which is a line that connects the values of the heights of the histogram, as shown in Figure 3.3. The selection of a bar chart or line plot (or other plots) may be a matter of personal choice or may be determined by historical practice of presenting similar data.

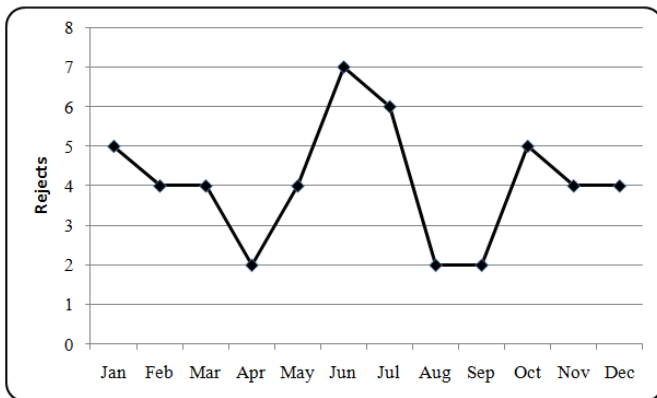


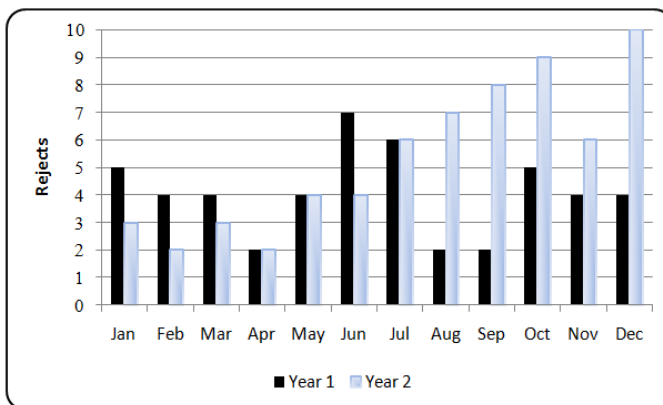
Figure 3.3. Line plot for monthly fuel rod rejects

Table 3.2 and Figures 3.2 and 3.3 are all missing some important data—the total number of fuel rods that were inspected. The effect of this deficiency may be illustrated by considering the data from 2 consecutive years, given in Table 3.3.

Table 3.3. Monthly rejects of fuel rods

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rejects, year 1	5	4	4	2	4	7	6	2	2	5	4	4
Rejects, year 2	3	2	3	2	4	4	6	7	8	9	6	10

The data in Table 3.3 are plotted in Figure 3.4, which is a multibar graph. The first impression from this graph is that performance in the second year of operation (“Year 2”) deteriorated, since the number of rejects increased. However, this pattern may have several plausible explanations without such negative implications. One explanation is that increased production in the second year resulted in a greater number of rejects but a smaller percentage of defective rods. Another explanation is that towards the middle of the second year, the production line instituted a more sensitive quality control process that classified more rods as defective. And there may be still other explanations.

**Figure 3.4. 2-year monthly fuel rod rejects, stacked**

Alternatively, we may plot the same data in sequence, as shown in Figure 3.5. The difference between Figures 3.4 and 3.5 reflects a difference in the purpose of the study. If we believe that there is a seasonal variation, Figure 3.4 is constructed to allow month-by-month comparison. If this is not the case, Figure 3.5 may be more appropriate.

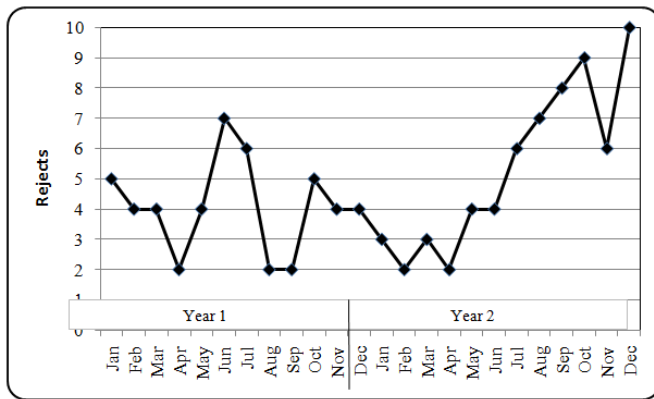


Figure 3.5. Line plot for 2-year monthly fuel rod rejects

3.6 Suggestions for constructing bar charts

The construction of bar charts, like that of the pie chart, has subjective and esthetic aspects. However, we must not let our artistic and creative impulses carry us too far. The following rules are provided:

1. The number of bars in the chart should be between 5 and 20. Having too many bars clutters the chart, while having too few bars may lessen the value of the chart.
2. If possible, the bars should have the same width, unless other considerations take precedence.
3. Bars are generally separated to emphasize the discrete nature of the grouping.
4. If possible, the distance between the bars should be constant, unless emphasis is needed.
5. One axis (typically the left) shows the actual frequency (also called frequency or absolute frequency). If both absolute frequency and relative frequency are shown, they could be reported using different vertical (y) axes.

Using Excel, bar charts may be constructed by following the steps listed below:

1. List the category labels in one column. Thus, for the data from Table 3.3, list the months (Jan through Dec) in cells A1..A12.

2. List the corresponding numerical values next to their labels in the adjacent columns (in this example, in cells B1..C12).
3. Select the cells that include the labels and the values. In this example, select cells A1..C12.
4. Click the **Chart** icon.
5. Click **Column**.
6. Select the button labeled **Column**.
7. Click **Finish** to display the bar chart.
8. Embellish the chart by selecting frame, color, font, shading, and other options as desired.

3.7 The histogram

histogram A *histogram* is a specialized type of bar chart; it is used primarily for displaying frequencies of a variable measured on an interval (metric, continuous) scale. The measurements are grouped into *class intervals*, usually of equal widths. The middle of each interval is called the *class marker*. For the purpose of display, all measurements falling into a particular class interval are considered to have the same value as the class marker. A histogram, by default, shows the relative frequencies (i.e., the fractions of the measurements that fall into each class interval) on the vertical axis. Since the variable is continuous, the spaces between the bars of the histogram are seldom shown.

Example 3.3. Water impurity measurements. In a test for impurities, 25 buckets of water were drawn randomly from a reservoir that was used to cool the reactor. Table 3.4 lists water impurities for each of the buckets, measured in parts per million (ppm), in ascending order of magnitude.

Table 3.4. Water impurities measured for 25 buckets, (ppm)

42.31	42.32	42.40	42.42	42.45
42.45	42.48	42.49	42.50	42.59
42.61	42.61	42.62	42.63	42.65
42.67	42.72	42.82	42.83	42.83
42.85	42.91	42.93	42.97	43.17

In Table 3.4, the smallest value is 42.31, and the largest value is 43.17. Now, suppose we decide upon a class interval width of 0.1 ppm and set up a scheme of class intervals that have “nice” (i.e., easily set and managed) class markers. For these particular data, we might decide on this set of class markers {42.3, 42.4, 42.5...43.2}. Given this choice, the first class interval contains all measurements less than or equal to 42.35 ppm; the second contains all those greater than 42.35 ppm and less than or equal to 42.45 ppm, and so on. In this fashion, every one of the 25 values can be placed unambiguously into one of the 10 intervals so defined. This process yields the frequencies that are the basis for the heights of the bars. In this example, two values out of 25, (or $2/25 = 0.08$) are associated with 42.3, four values out of 25, (or $4/25 = 0.16$) are associated with 42.4, and so on, and one value out of 25, (or $1/25 = 0.04$) is associated with 43.2. With these associations, we can proceed to construct the histogram, as shown in Figure 3.6.

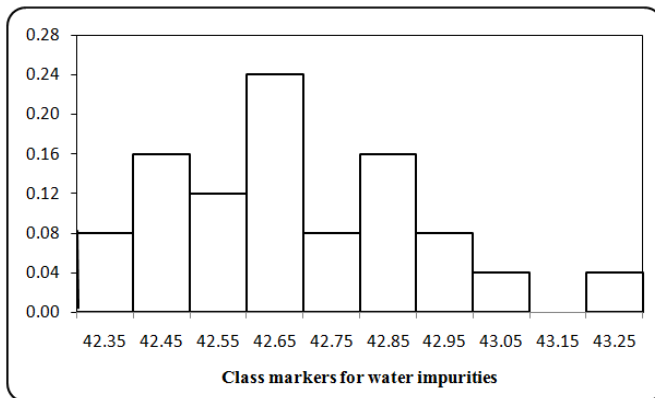


Figure 3.6. Histogram for data in Table 3.4

Histograms have several shortcomings that warrant our attention. Some detail is lost by assigning individual values the value of their class marker. Additionally, the number of class intervals, as well as the determination of the class markers, may be arbitrary. Consequently, the analysis and the interpretation of the histograms may differ if constructed by different analysts.

When the number of observations is small, plotting these values in a histogram may be unwarranted. Thus, in Example 3.3, we may plot the data in a so-called *dot plot* that preserves the recorded value, as shown in Figure 3.7. The choice between a dot plot and a histogram, however, depends on the subject matter, management preference, and historical and traditional practice.

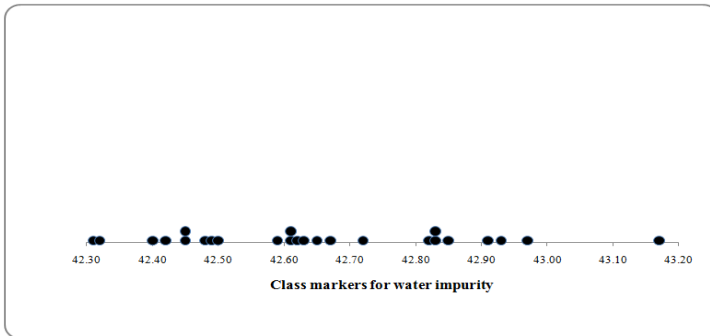


Figure 3.7. Dot plot for data in Table 3.4

3.8 Suggestions for constructing histograms

The rules for constructing a histogram are essentially the same as the rules for constructing a bar chart:

1. The number of bars in the histogram should be between 5 and 20. Having too many bars clutters the chart, while having too few bars may lessen the value of the histogram.
2. Be sure that every data point belongs to one and only one class interval.
3. Class intervals should have the same width, unless other considerations take precedence.
4. One axis (typically the left) shows the relative frequency. If both frequency and relative frequency are shown, they are reported using different vertical (y) axes.
5. Bars should be drawn without spaces between them.

Excel may be used to produce a histogram in several ways, and we present one of them. We start by building a column chart, and instead of group labels, we use class markers. Class markers, however, must be entered as alphanumeric characters rather than pure numeric. This can be accomplished in Excel by placing an apostrophe before each class's numeric value, such as '42.5. Then follow these steps:

1. List the class markers in alphanumeric form in one column. Thus, in the data from Table 3.4, where we have 10 classes, enter the values in cells A1..A10 as '42.3, '42.4, ... '43.2.

2. List the corresponding fraction (proportion) of the values for each class next to their class marker in the adjacent columns (in this example, in cells B1..B10). Since we have 25 values, the fraction entered in cell B1 is $2/25 = 0.08$.
3. Select the cells that include the class markers and their associated values. In this example, select cells A1..C10.
4. Click on the **Chart** icon.
5. Click on **Column**.
6. Click on **2D Column**.
7. Click on **Finish** to display a bar chart.

Since a histogram usually contains no gaps between the class markers, we continue:

8. Right-click any column on the bar chart.
9. Click on **Format data series**.
10. Click on **Gap width**.
11. Move the slider to **No gap**.
12. Embellish the chart by selecting frame, color, font, shading, and other options as desired.

3.9 The box plot

box plot A *box plot* is a graphical display designed to represent a set of data in a relatively small amount of space and yet give a picture of the extent of the dataset, as well as some of its distributional features. There are many versions of the box plot, with various levels of detail.

The five fundamental quantities that determine a box plot are obtained as shown below. These fundamental quantities are illustrated using the water impurity data in Table 3.4.

minimum The *minimum* (min) is the smallest value in the dataset.
The minimum of the 25 water impurities values is 42.31.

maximum The *maximum* (max) is the largest value in the dataset.
The maximum of the 25 water impurities values is 43.17.

Median
second quartile The **median** (**second quartile**, 50th percentile) is the middle value after the values are arranged in ascending order of magnitude.

The median divides the dataset into two equal-sized groups. As defined in Section 2.3:

- The median is the middle value if the number of values is odd.
- The median is the average of the two middle values if the number of values is even.

The median of the 25 water impurities values is 42.62.

lower quartile (LQ) The **lower quartile (LQ)** is the median of the group containing the values below the dataset's median.

For the 25 water impurities values, $LQ = (42.45 + 42.48) / 2 = 42.465$.

For this dataset, the lower quartile equals the 25th percentile or first quartile (see Section 2.6). In general, the lower quartile may only approximate the 25th percentile of the dataset.

- From the definition of percentile in Section 2.6, it can be shown that the lower quartile and 25th percentile are always equal if the number of points in the dataset $n = 4k + 1$, for some integer k . Table 3.4 satisfies this condition because $25 = (4)(6) + 1$.

upper quartile (UQ) The **upper quartile (UQ)** is the median of the group containing the values above the dataset's median.

For the 25 water impurities values, $UQ = (42.83 + 42.83) / 2 = 42.83$.

In parallel with the lower quartile, the upper quartile is equal to the 75th percentile, or third quartile, for this dataset. Also, in general, the upper quartile may only approximate the 75th percentile of the dataset.

Figure 3.8 illustrates a box plot summarizing the water impurity data in Table 3.4.

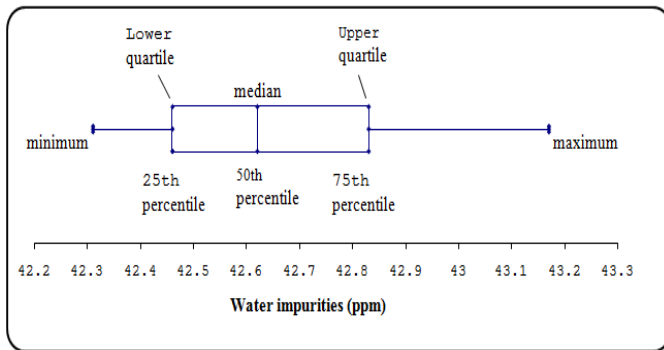


Figure 3.8. Box plot of 25 water impurity measurements

- box** The lower and upper quartiles in Figure 3.8 are enclosed in a **box**. That box also shows the location of the median by a vertical line.
- whiskers** The two protruding lines outside the box are called **whiskers**. In a simple version of a box plot, one whisker is drawn as a line from the minimum to the lower quartile, and the second whisker is drawn as a line from the upper quartile to the maximum.
- box and whisker** A box plot is sometimes called a **box and whisker** plot. Box plots can also be erected vertically with the whiskers extending upward and downward.

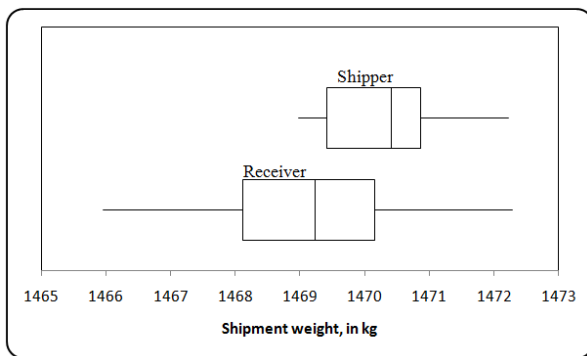
Box plots are of special value for a visual comparison of two or more sets of data with respect to the same variable.

Example 3.4. Shipper and receiver reported weights. Table 3.5 lists shipping weights of uranium hexafluoride (UF_6) cylinders reported by both shipper and receiver for 10 shipments and their weight differences. This table also presents selected statistics. Note, however, that the last four statistics in the table are not relevant to the immediate discussion but will be introduced and discussed later in the chapter.

Table 3.5. Shipper's and receiver's weights of UF₆ cylinders, in kilograms

Shipment	Shipper	Receiver	Difference
1	1471.22	1468.12	3.10
2	1470.98	1469.52	1.46
3	1470.82	1469.22	1.60
4	1470.46	1469.26	1.20
5	1469.42	1465.96	3.46
6	1468.98	1470.80	-1.82
7	1469.10	1467.89	1.21
8	1470.22	1472.28	-2.06
9	1470.86	1469.02	1.84
10	1470.38	1470.16	0.22
Minimum	1468.98	1465.96	-2.06
Lower quartile	1469.42	1468.12	0.22
Median: 50 th percentile	1470.42	1469.24	1.34
Upper quartile	1470.86	1470.16	1.84
Maximum	1471.22	1472.28	3.46
Range	2.24	6.32	5.52
IQR	1.44	2.04	1.62
Lower fence	1467.26	1465.06	-2.21
Upper fence	1473.02	1473.22	4.27

Figure 3.9 shows box plots of shipper's and receiver's data. It is easy to see from that figure that the receiver is reporting fewer kilograms of material than the shipper is claiming. In addition, the receiver's values are more variable than the shipper's, as reflected in the receiver's much larger range. The box plot clearly conveys these impressions.

**Figure 3.9. Box plots of shipper's and receiver's weights of UF₆ cylinders**

Box plots or, for that matter, any plot or analysis must be constructed with careful attention to the content. For example, in Table 3.5, it is useful to examine the differences between the shipper and receiver entries. These differences, reported in the last column of Table 3.5, are plotted as a box plot in Figure 3.10. That figure shows that, in a majority of the cases, the receiver claims to have received less than the shipper claims to have shipped. However, if we wish to discover whether these differences are related to other factors, further statistical investigation may be warranted. For example, a regression analysis (Chapter 18) may be run to investigate whether these differences are related to the size of the shipment, the shipping date, or the distance the shipment traveled.

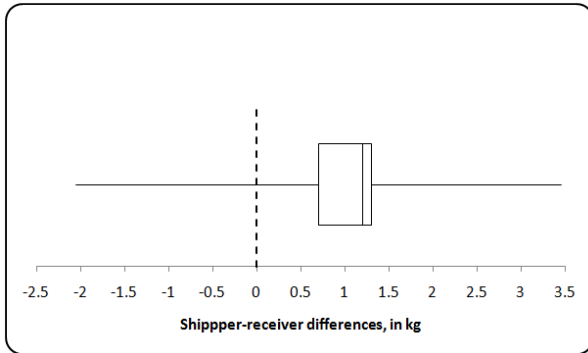


Figure 3.10. Box plot of shipper-receiver differences

If two or more plot boxes are shown below (or next) to each other, the thickness (width) of the boxes may be drawn proportionally to reflect different sample sizes.

Embellishments of the box and whiskers in a box plot are often added to make the presentation more informative. Such embellishments require additional definitions. These definitions are associated with various measures of the variability of the data.

interquartile range (IQR) The *interquartile range (IQR)* is the length of the box, which is the difference between the upper and the lower quartiles. For Table 3.4, *IQR* is calculated as $42.83 - 42.465 = 0.365$. For the Shipper, Receiver, and Difference columns of Table 3.5, *IQR* values are calculated as

$$1470.86 - 1469.42 = 1.44, 1470.16 - 1468.16 = 2.04, \\ \text{and } 1.84 - 0.22 = 1.62, \text{ respectively.}$$

fence The *fences* are defined, relative to the box plot, as follows:

- lower fence** The **lower fence** is equal to $LQ - 1.5(IQR)$. For Table 3.4, the lower fence is calculated as $LF = 2.465 - 1.5(0.365) = 41.918$. For Table 3.5, the three lower fences are 1467.26, 1465.06, and -2.21.
- upper fence** The **upper fence** is equal to $UQ + 1.5(IQR)$. For Table 3.4, the upper fence is calculated as $UF = 42.83 + 1.5(0.365) = 43.378$. For Table 3.5, the three lower fences are 1473.02, 1473.22, and 4.27.
- outlier** An **outlier** is defined in the context of the box plot as a data value that is smaller than the lower fence or larger than the upper fence. Outliers are marked with a special symbol, such as a circle. Chapter 26 gives other definitions of outliers, and the methodology for their identification.

Box plots may be embellished in various ways. Embellished box plots typically have different whisker construction. For example, one such embellished construction draws one whisker from the left side of the box to the smallest value that is not an outlier; the second whisker is drawn from the right side of the box to the largest value that is not an outlier.

Owing to its particular power as a data analysis tool, the box plot can be embellished even further. See Chambers, et al. (1983), p. 21, for an extended discussion of the box plot.

3.10 Suggestions for constructing box plots

The rules for constructing a box plot vary according to the level of detail that the data deserve:

1. Calculate the minimum, maximum, median, and lower and upper quartiles from the data.
2. Draw a rectangular box, with vertical ends at the lower and upper quartiles.
3. Draw a vertical line inside the box at the median and parallel to the hinges.
4. Calculate the IQR , the difference between the upper and lower quartiles.
5. Calculate the upper fence as the upper quartile plus $1.5(IQR)$ and the lower fence as the lower quartile less $1.5(IQR)$.
6. Draw a line (whisker) from the upper quartile to the maximum and, similarly, from the lower quartile to the minimum.

- Note that there are competing versions of the box plot e.g. a vertical oriented box, rather than a horizontal box used in this chapter. Should your report contain a box plot, include some explanation of its features.
7. More than one box plot may be drawn on the same chart.
 8. Different box plots on the same chart may have different thicknesses (widths) to reflect another variable, such as count or volume.

Finally, be aware that Excel does not offer a direct routine for plotting box plots. In contrast, box plotting is available in many, if not most, statistical packages.

3.11 The stem-and-leaf display

Histograms display the structure of the data by grouping them into intervals. Such groupings may lead to a loss of detail. For example, in Table 3.4, the water impurities of 42.67 and 42.72 are both assigned the value of 42.7. Loss of detail may be detrimental to a sensitive analysis.

stem-and-leaf display A *stem-and-leaf display* tends to overcome this obstacle by providing a short-hand notation that shows almost all the details of the data. The following simplified example explains this display.

Example 3.5. Stem-and-leaf display. Consider the numbers 5231, 5234, 5234, and 5236. Their common leading digits 523 are called the stem. Each trailing digit (1, 4, 4, and 6) is called a leaf. The stem-and-leaf notation for these four numbers is 523|1446. Similarly, the stem-and-leaf notation 20|13477888 represents the eight numbers 201, 203, 204, 207, 207, 208, 208, and 208.

In this notation, the stem is written to the left of the vertical bar. Most datasets have more than one stem, and those stems are written in ascending order. The leaves, written to the right of the vertical line, are also written in ascending order.

In a similar manner, Figure 3.11 lists the data from Table 3.4 in a stem-and-leaf display. Note that not only are all the data retrievable, but the stacking of the leaves generates a histogram that can be easily read by turning the page 90 degrees.

<u>Stem Leaf</u>
42.3 12
42.4 025589
42.5 09
42.6 112357
42.7 2
42.8 2335
42.9 137
43.0
43.1 7

Figure 3.11. Stem-and-leaf display for data in Table 3.4

Figure 3.12 is an embellished display that adds two features. The first feature omits the decimal point in the stem and replaces it with two equivalent statements. The first statement, “Unit = 0.01,” indicates that each leaf is equivalent to a multiple of 0.01. The second statement indicates, in a simpler manner, that the first leaf of the first stem is 42.31 and is represented by 4231. Clearly, with either of these two statements, the display is fully defined.

<u>Depth</u>	<u>Stem Leaf</u>
2	423 12
8	424 025589
10	425 09
(6)	426 112357
9	427 2
8	428 2335
4	429 137
1	430
1	431 7

Unit = 0.01 423|1 represents 42.31

Figure 3.12. Embellished stem-and-leaf display

A third feature of the embellished Figure 3.12 is that it counts and accumulates the number of leaves in each row separately for each side of the distribution about the median. The counts in the accumulated column characterize the *depth* (i.e., the total number of leaves in the rows counting from either end of the display). Thus, there are two leaves in the first row and six in the second for an accumulated total depth of $6 + 2 = 8$. The row that includes the median shows the frequency (in parentheses) of that row alone, rather than the accumulated count. Also note that the total number of values in the dataset is equal to the sum of the counts for the median row, the row above the median, and the row below the median. For Figure 3.11, that count is $6 + 9 + 10 = 25$.

The value of the stem-and-leaf display may not be evident for a small dataset such as is found in Table 3.4. The value of this display is more obvious when dealing with a large dataset. To demonstrate the display, we used combined data from Bowen and Bennett (1988), p. 13 of weights (in kilograms) of 150 uranium ingots. The raw data are not reproduced here, but the stem-and-leaf display is constructed and shown in Figure 3.13. Although not required for the construction of a stem-and leaf display, other statistics are presented at the bottom of Figure 3.13 to allow the construction of a box plot for the same data.

Depth	Unit = 0.1	418 5 represents 418.5
1	418 5	
3	419 27	
4	420 4	
11	421 0237889	
15	422 1445	
28	423 1122333566679	
47	424 0224455667888889999	
67	425 01122233444577778999	
(18)	426 011223333456667889	
65	427 0122333333445578899999	
43	428 0000022334568889	
7	429 012345677789	
15	430 00359	
10	431 012367	
4	432 234	
1	433	
1	434 2	
Minimum	418.5	
Maximum	434.2	
Lower quartile	424.7	
Median	426.3	
Upper quartile	428.2	
<i>IQR</i>	3.5	
Lower fence	419.45	
Upper fence	433.45	

Figure 3.13. Stem-and-leaf display for 150 uranium ingot weights

As Figure 3.13 illustrates, the raw data can be completely recaptured from the stem-and-leaf display. That display also gives a histogram, although the class markers are restricted to the leading digits.

An interesting feature of the stem-and-leaf display is that when turned on its side, the empirical distribution of the data may be seen. In Figure 3.12, that distribution appears as a mound-shaped distribution.

There is much more to the stem-and-leaf display than presented here. For further discussion of stem-and-leaf displays and various embellishments, see Tukey (1977, p. 8) or Velleman and Hoaglin (1981).

3.12 Suggestions for constructing a stem-and-leaf display

As for the box plot, the rules and the embellishments of the stem-and-leaf display vary according to the taste and the sophistication of the developer and the audience. The basic rules are as follows:

1. Use a stem-and-leaf display for a reasonably large database. A display for a small database may not yield additional information.
2. Arrange the data in ascending order.
3. Determine the structure of the stems and the leaves.
4. Arrange the stems in ascending order from top to bottom and the leaves in ascending order from left to right.
5. Count the leaves for each stem.
6. Determine the number of leaves in the row that includes the median. Place this number in parentheses to the left of that stem.
7. For each stem below the median, record, to the left of the stem, the accumulated number of leaves at and below that stem.
8. For each stem above the median, record, to the left of the stem, the accumulated number of leaves at and above that stem.
9. If the decimal place is omitted, indicate how the leaves and the stem represent the data.

4

Basics of probability

4.1 What to look for in Chapter 4

This chapter introduces probability, which is the foundation for the statistical topics covered in the remainder of this book. These topics concepts and topics include:

- *probability, § 4.2*
- *experiment, § 4.3*
- *outcome, § 4.3*
- *event, § 4.3*
- *sample space, § 4.3*
- *Venn diagram, § 4.4*
- *complementary events, § 4.4*
- *mutually exclusive events, § 4.4*
- *union of events, § 4.4*
- *intersection of events, § 4.4*
- *basic rules of probability, § 4.5*
- *marginal probability and joint probability, § 4.6*
- *conditional probability, § 4.7*
- *independent events, § 4.7*
- *Bayes' theorem, § 4.8*
- *probability quantification, § 4.9*

4.2 The concept of probability

Many phenomena and events in the real world are uncertain. For example, we may be uncertain whether it will rain tomorrow, who will be elected the next president, and whether a nuclear power plant will suffer a core damage accident in its lifetime. All these situations have elements of uncertainty, some larger than others. Uncertainty about phenomena or event outcomes is often described by such terms such as “unlikely,” “possible,” “likely,” and “probable.” These expressions are often preceded by adverbs such as “very,” “highly,” and “extremely.”

probability **Probability** may be characterized as a numerical measure of uncertainty. The concept of probability is illustrated by the following example.

All nuclear power plants have emergency diesel generators (EDGs) that are designed to provide power in the event of a loss of offsite power. A key parameter in the risk assessment of such an event is the probability P_{start} that an EDG will start when activated. Because this book is about the application of statistics to problems in nuclear regulation rather than a theoretical treatise on statistics, we will consider P_{start} for a specific EDG operating in a specific environment at a specific plant. This means that P_{start} , like all probabilities, depends on the specific conditions for which it is defined. While these conditions may vary from one EDG activation to another, we assume that they are sufficiently constant so that the operating conditions of the EDG are identical for each activation.

In common English usage, probability is synonymous with chance or likelihood. In this book, probability will always mean a numerical measure of the uncertainty associated with the outcome of an uncertain or unpredictable event. Thus, P_{start} is a number that is a measure of the chance that the specific EDG will start when activated.

Although it may not be immediately apparent, this definition of the probability P_{start} is directly analogous to other characteristics of the specific EDG such as its weight, w . Here, w is a measure of the force of attraction between the mass of the EDG and the Earth, and whose numerical value must be determined by measurement. At this point, one might object that the analogy between weight and probability breaks down because a single measurement is sufficient to determine a value for w , while P_{start} can only be estimated. However, a single measurement can never determine the exact value of w . For example, if a scale accurate to a kilogram is used to weigh the EDG, it is plausible that repeated measurements (weightings) may all yield the same value, w_0 , for w . In such a situation, we might assert that $w = w_0$, but this would not be correct. In fact, all that could be claimed is that $w_0 - 0.5 \leq w \leq w_0 + 0.5$. If a more accurate scale is used (e.g., one accurate to a gram or a milligram), then no two measurements are likely to be the same. Thus, regardless of the accuracy of the scale, no single

measurement can determine the exact value of w . (In practice, only one measurement may be necessary if it is sufficiently accurate for the application. However, the exact value of w will not be known.)

A continuation of the above discussion serves to clarify the analogy between weight and probability. To obtain a value for w with less than half a kilogram uncertainty, a number of measurements might be made and the average used. However, this average would still be an approximation for w and not its exact value. Thus, no matter how accurate a scale is, or how many measurements are made, it is impossible to obtain an exact value for w . This is the same situation as determining an exact value for P_{start} . No matter how many measurements are made, it is impossible to obtain an exact value for P_{start} ; it can only be estimated.

It is important to distinguish between the definition of P_{start} and a determination of its value. In practice, probabilities are quantified by using some combination of modeling, data, and judgment. The result can be an exact number if it is based on a completely specified model or an approximation if it is based on data or judgment. Section 4.9 will discuss approaches to quantifying probability.

4.3 Sample spaces and events

<i>experiment</i>	An activity or investigation for which the results are uncertain is called an <i>experiment</i> . Examples of experiments are tossing a coin, rolling a die, counting failures over time, and measuring time to failure of an electric bulb. The result of one execution of
<i>outcome</i>	the experiment is referred to as an <i>outcome</i> . Repetitions of a defined experiment would not be expected to produce the same outcomes due to the uncertainty associated with the process. The set of all possible outcomes of an experiment is defined as the
<i>sample space</i>	<i>sample space</i> . We denote a sample space by S .

Example 4.1. Tossing a coin. In the experiment of a coin tossing, we have two outcomes: heads (H) and tails (T). The sample space $S = \{H, T\}$ consists of these two outcomes.

Example 4.2. Rolling a die. In rolling a die, the outcomes are the six faces of the die. The sample space $S = \{1, 2, 3, 4, 5, 6\}$ is the set of the six faces.

Example 4.3. Rolling a die (continued). Let the event E denote the roll of an even number. Then $E = \{2, 4, 6\}$ consists of these three outcomes.

<i>event</i>	Sample spaces can contain discrete points (such as pass, fail) or points in a continuum (such as measurement of time to failure). An event E is a specified set of possible outcomes in a sample space S and is written as $E \subset S$, where the symbol \subset denotes “subset of.”
<i>empty set</i> <i>null set</i>	A special set is the set without any elements. It is called the empty set or null set and is denoted by \emptyset .
<i>elementary event</i> <i>compound event</i>	If an event cannot be decomposed into smaller parts, we call it an elementary event . Otherwise, we call it a compound event . All of the outcomes in a sample space such as the outcomes described in Example 4.1 and 4.2 are elementary events. Most events of interest in practical situations are compound events, formed by some composition of two or more elementary events, such as receiving an A in math and a B in chemistry.

4.4 Basic set theory

<i>sample points</i> <i>universal set</i>	A systematic development of probability sometimes uses the mathematical theory of sets. Composition of events can occur through the union, intersection, or complement of events, or through some combination of these. The outcomes in a sample space are also called sample points . In the language of set theory, a sample space is called the universal set .
<i>union</i>	For two events, E_1 and E_2 , in a sample space S , the union of E_1 and E_2 , denoted by $E_1 \cup E_2$, is defined to be the event containing all sample points in E_1 or E_2 or both. Thus, the union is the event that either E_1 or E_2 or both E_1 and E_2 occur.
<i>intersection</i>	For two events, E_1 and E_2 , in a sample space S , the intersection of E_1 and E_2 , denoted by $E_1 \cap E_2$, is defined to be the event containing all sample points that are in both E_1 and E_2 . Thus, the intersection is the event that both E_1 and E_2 occur.
➤	The intersection symbol \cap is often omitted (i.e., $A \cap B$ is written as AB).

- complement** The **complement of an event** E in a sample space S is the set of all sample points in S that are not in E . The complement of E is denoted in this book by E^c but is often denoted by the symbol \bar{E} (read E-bar) or E' (read E-prime).
- mutually exclusive** Two events, E_1 and E_2 , in a sample space S are said to be **mutually exclusive** if the event $E_1 \cap E_2$ contains no outcomes in the sample space S . That is, the intersection of the two events is \emptyset . Mutually exclusive events are also referred to as **disjoint events**.
- disjoint events** Three or more events are said to be mutually exclusive, or **disjoint**, if each pair of events is mutually exclusive. In other words, no two events can occur together.
- Venn diagram** Figure 4.1 shows a schematic picture, called a **Venn diagram**, of 10 outcomes and 3 events. In this example, the event E_1 contains three outcomes, event E_2 contains five outcomes, and event E_3 contains three outcomes. The intersection of E_1 and E_2 contains one point, and E_3 is disjoint from E_1 and from E_2 .

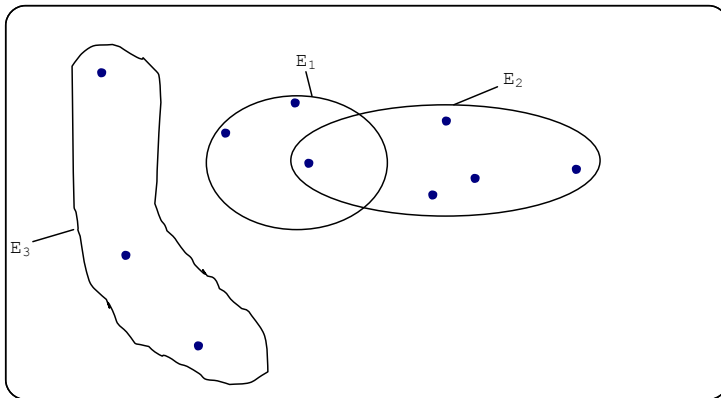


Figure 4.1. Venn diagram showing 10 outcomes and 3 events

Example 4.4. Three Sets. Consider the sample space illustrated in Figure 4.2. The three circles, denoted by A , B , and C , correspond to three events. The complement of their union, denoted by $VIII$, is also an event. As shown, the three circles and their complement can be decomposed into eight mutually exclusive events denoted by I , II , III , IV , V , VI , VII , and $VIII$.

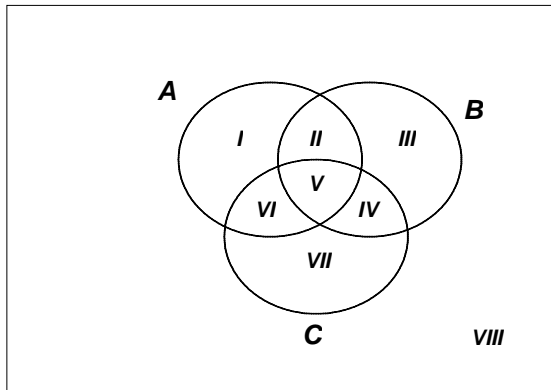


Figure 4.2. Decomposition of three sets

The relationships of Table 4.1 are shown below.

Table 4.1. List of sets in Figure 4.2

$A = I \cup II \cup V \cup VI$	$I = AB^cC^c$
$B = II \cup III \cup IV \cup V$	$II = ABC^c$
$C = IV \cup V \cup VI \cup VII$	$III = A^cBC^c$
$A^c = III \cup IV \cup VII \cup VIII$	$IV = A^cBC$
$B^c = I \cup VI \cup VII \cup VIII$	$V = ABC$
$C^c = I \cup II \cup III \cup VIII$	$VI = AB^cC$
	$VII = A^cB^cC$
	$VIII = A^cB^cC^c$

Example 4.5. Valve failures. Suppose that 10 valves in a nuclear plant are actuated. Let A_i denote the event that valve i operates successfully and F_i denote the event that valve i fails, $i = 1, 2, \dots, 10$. Then the sample space is $S = \{A_1, F_1, \dots, A_{10}, F_{10}\}$.

Example 4.6. Time to failure. Suppose that 10 pumps in a nuclear plant are required to operate continuously over a 1-year period (8,760 hours). If pump i , $i = 1, 2, \dots, 10$ operates successfully for 1 year, denote this event by B_i ; if pump i fails after t_i hours, denote this event by $F_i(t_i)$ for $0 < t_i \leq 8760$ hours. The sample space for this example is $S = \{B_1, F_1(t_1), \dots, B_{10}, F_{10}(t_{10})\}$.

Set operations can be combined in many ways. Table 4.2 lists the most common combinations and their properties.

Table 4.2. Properties of set operations

Complement laws	$A \cup A^c = S$ $(A^c)^c = A$ $A \cap A^c = \emptyset$
Commutative laws	$A \cup B = B \cup A$ $A \cap B = B \cap A$
DeMorgan's laws	$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$
Associative laws	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
Distributive laws	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Null set properties	$A \cup \emptyset = A$ $A \cap \emptyset = \emptyset$
Universal set properties	$A \cup S = S$ $A \cap S = A$ $A \cup A^c = S$

4.5 Basic rules and principles of probability

probability model

In Section 4.2, probability is characterized as a numerical measure of the uncertainty associated with the outcome of an uncertain event. Although the events we are concerned with all involve physical objects and phenomena, we will represent and analyze them using a **probability model**. This is completely analogous to the use of mathematical constructs such as Newton's laws and Einstein's theory of relativity to describe the behavior of physical objects and phenomena in the real world.

probability function axioms of probability

A probability model consists of a sample space, S , together with a **probability function** associated with each event E in S , written $Pr\{E\}$, such that the following conditions, called the **axioms of probability**, apply:

Axiom 1. $0 \leq Pr\{E\} \leq 1$ for every E in S . (4.1)

Axiom 2. $Pr\{E_1 \cup E_2 \cup \dots\} = Pr\{E_1\} + Pr\{E_2\} + \dots$ for any set of disjoint events $\{E_1, E_2, \dots\}$ in S . (4.2)

Axiom 3. $Pr\{S\} = 1$. (4.3)

probability function A **probability function** specifies how the probability is distributed over various subsets E of a sample space S . The axioms of probability imply a number of rules of probability that provide additional properties of a probability function.

An important special case of Equation (4.2) occurs when the union of the disjoint events E_1, E_2, \dots is the entire sample space S (i.e., $S = E_1 \cup E_2 \cup \dots$ for any set of disjoint events $\{E_1, E_2, \dots\}$ in S). It follows from Equations (4.2) and (4.3) that:

$$Pr\{E_1\} + Pr\{E_2\} + \dots = 1 \quad (4.4)$$

partition In such a case, the set $\{E_1, E_2, \dots\}$ is called a **partition** of the sample space.

If E is any event in a sample space S and $\{S_1, S_2, \dots\}$ is a partition of S , then $Pr\{E\}$ can be written as:

$$Pr\{E\} = Pr\{E \cap S_1\} + Pr\{E \cap S_2\} + \dots \quad (4.5)$$

Equation (4.5) is based on a generalization of the second distributive law in Table 4.2 whereby:

$$E \cap (S_1 \cup S_2 \cup \dots) = (E \cap S_1) \cup (E \cap S_2) \cup \dots \quad (4.6)$$

Because $\{S_1, S_2, \dots\}$ is a partition of S , the left side of Equation (4.6) can be written as:

$$E \cap (S_1 \cup S_2 \cup \dots) = E \cap S = E \quad (4.7)$$

Because the sets on the right side of Equation (4.6) are disjoint, Equation (4.5) follows from the application of Equation (4.2) to Equation (4.6).

empty set The probability of an impossible event (the **empty set** or **null set**, denoted by \emptyset) is zero. Symbolically, $Pr\{\emptyset\} = 0$. This result follows from Equation (4.2) because $E \cup \emptyset = E$ for any event in S .

null set

The probability of the complement of event E is given by:

$$Pr\{E^c\} = 1 - Pr\{E\} \quad (4.8)$$

Equation (4.8) follows from Equations (4.2) and (4.3) and the relation $E \cup E^c = S$ for any event E .

In general, the probability of the union of any two events is given by:

$$Pr\{E_1 \cup E_2\} = Pr\{E_1\} + Pr\{E_2\} - Pr\{E_1 \cap E_2\} \quad (4.9)$$

A visual proof of Equation (4.9) follows from consideration of sets A and B in Figure 4.2. From the figure, the probability of the intersection of A and B is double-counted by summing their probabilities and must therefore be subtracted to get the probability of their union.

If E_1 and E_2 are mutually exclusive, then:

$$\begin{aligned} Pr\{E_1 \cap E_2\} &= Pr\{\emptyset\} = 0, \text{ and} \\ Pr\{E_1 \cup E_2\} &= Pr\{E_1\} + Pr\{E_2\} \end{aligned} \quad (4.10)$$

addition rule for probabilities The second relation in Equation (4.10) is the second axiom of probability for two events. Equation (4.2) is sometimes referred to as the **addition rule for probabilities**.

For three events,

$$\begin{aligned} Pr\{E_1 \cup E_2 \cup E_3\} &= Pr\{E_1\} + Pr\{E_2\} + Pr\{E_3\} \\ &\quad - Pr\{E_1 \cap E_2\} - Pr\{E_1 \cap E_3\} - Pr\{E_2 \cap E_3\} \\ &\quad + Pr\{E_1 \cap E_2 \cap E_3\} \end{aligned} \quad (4.11)$$

inclusion-exclusion principle

This result follows from the decomposition in Figure 4.2 in a similar manner to the derivation of Equation (4.9). This rule is also referred to as the **inclusion-exclusion principle** and can be generalized to n events. It is widely used in probabilistic risk assessment (PRA) to calculate the probability of an “or” gate (a union of events) in a fault tree used for accident sequence quantification (NRC (2003), p. A-3).

rare event approximation The inclusion-exclusion principle also provides useful upper and lower bounds on the probability of the union of n events that are not mutually exclusive. One such upper bound, referred to as the *rare event approximation*, is:

$$\begin{aligned} Pr\{E_1 \cup E_2 \cup \dots \cup E_n\} &\leq Pr\{E_1\} + \\ &Pr\{E_2\} + \dots + Pr\{E_n\} \end{aligned} \quad (4.12)$$

If the n events are mutually exclusive, the error in the rare event approximation is zero. Otherwise, the rare event approximation should be used only when the probabilities of the n events are all very small (NRC, 2003, p. A-3).

An approximation of the error in the rare event approximation is $[n(n-1)/2] \max [Pr\{E_i\}]$, which is valid regardless of the relationships between the events (NRC (2003), p. A-3). This approximation is frequently used in accident sequence quantification.

4.6 Marginal and joint probabilities

marginal probability The probability that an event will occur, regardless of whether other events occur, is called the *marginal probability* of that event. For example, the probability $Pr\{R\}$ of a car jumping the red light at a given time at a given intersection, regardless of whether a pedestrian is in the crosswalk, is a marginal probability. Similarly, the probability $Pr\{P\}$ that a pedestrian is in the crosswalk at a given time, regardless of the traffic, is also a marginal probability.

joint event
joint probability On the other hand, many situations arise where the outcomes are defined by two or more events occurring simultaneously. For example, a catastrophic event takes place when, at a given time and place, a car runs a red light and a pedestrian is in the crosswalk. The simultaneous occurrence of two or more events (the intersection of events) is called a *joint event*, and its probability is called a *joint probability*. Using the notation of Section 4.5, the joint probability of both events E_1 and E_2 occurring simultaneously is written as $Pr\{E_1 \cap E_2\}$ or $Pr\{E_1 E_2\}$.

statistically independent events

A very important special case of a joint event is that of statistical independence. Two events E_1 and E_2 are said to be **statistically independent events** if the joint probability of E_1 and E_2 equals the product of their marginal probabilities of E_1 and E_2 . Mathematically,

$$Pr\{E_1 \cap E_2\} = (Pr\{E_1\})(Pr\{E_2\}) \tag{4.13}$$

Example 4.7. Rolling two dice. A die is said to be fair if the probability of any face coming up is the same for each face of the die. When rolling two fair dice, there are 36 outcomes, as listed in Table 4.3. Because each die has six faces, it follows from Axioms 2 and 3 that the probability of any face is equal to $1/6$. If it is further assumed that the rolls are physically independent (i.e., the separate outcomes of the two rolls do not influence each other), then the separate outcomes are statistically independent. It follows from Equation (4.11) that the probability of each outcome listed below is equal to $1/36$. (As a check, note that the sum of the probabilities of all the outcomes is $36(1/36) = 1$.) Table 4.4 lists the probabilities of the 36 outcomes in Table 4.3.

Table 4.3. Sample space of rolling two dice

Die 1	Die 2					
	1	2	3	4	5	6
1	{1,1}	{1,2}	{1,3}	{1,4}	{1,5}	{1,6}
2	{2,1}	{2,2}	{2,3}	{2,4}	{2,5}	{2,6}
3	{3,1}	{3,2}	{3,3}	{3,4}	{3,5}	{3,6}
4	{4,1}	{4,2}	{4,3}	{4,4}	{4,5}	{4,6}
5	{5,1}	{5,2}	{5,3}	{5,4}	{5,5}	{5,6}
6	{6,1}	{6,2}	{6,3}	{6,4}	{6,5}	{6,6}

The outcomes in Example 4.7 can be characterized in terms of two sample spaces, one describing the outcome of rolling the first die (S_1) and the other describing the outcome of rolling the second die (S_2). The sample space for the joint outcome can be written as $T = (S_1, S_2)$, where T consists of the 36 sample points listed in Table 4.3. Table 4.4 gives the probability distribution of T , along with the marginal probability distributions for each of the two dice. For each outcome of Die 1 (corresponding to a row of Table 4.4), the marginal probability is the sum of the joint probabilities for that outcome (i.e., the sum of the

probabilities over all possible outcomes of Die 2 [corresponding to the columns of the table]). The marginal probabilities for Die 1 are listed in the last column in the table. Note that the marginal probability distribution for Die 1 is identical to the probability distribution for Die 1 (i.e., the probability distribution of its sample space, S_1). Similarly, the marginal probabilities for Die 2 are the sums over the rows of the table and are listed in the last row of the table.

Table 4.4. Joint and marginal probabilities for two fair dice

Die 1	Die 2						Marginal Probability
	1	2	3	4	5	6	
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Marginal Probability	1/6	1/6	1/6	1/6	1/6	1/6	1

Example 4.7 illustrates the case where S_1 and S_2 are independent sample spaces. The joint probabilities are then probabilities of statistically independent events and, from Equation (4.10), are equal to the products of the marginal probabilities. However, S_1 and S_2 are not always independent. In such cases, the joint probabilities are probabilities of dependent events. This means that the probability of an outcome of the sample space S_2 depends on the outcome of the sample space S_1 . This is illustrated by the following example.

Example 4.8. Weather on successive days. Weather tends to be persistent. This means that a day is more likely to be sunny than cloudy or rainy if the previous day has been sunny. Consider the possible weather combinations on two successive days at some location. Let $D_i = \{S_i, C_i, R_i\}$ be the sample space of weather on either the first ($i = 1$) or second ($i = 2$) day, where the outcomes are sunny (S_i), cloudy (C_i), or rainy (R_i). Suppose that Table 4.5 gives the joint and marginal probabilities for D_1 and D_2 .

Table 4.5. Joint and marginal probabilities for weather on successive days

	Day 2			Marginal probability (Day 1)
Day 1	Sunny (S_2)	Cloudy (C_2)	Rainy (R_2)	
Sunny (S_1)	0.40	0.05	0.05	0.50
Cloudy (C_1)	0.05	0.15	0.10	0.30
Rainy (R_1)	0.05	0.10	0.05	0.20
Marginal probability (Day 2)	0.50	0.30	0.20	1.00

unconditional probability

As in Example 4.7, the marginal probabilities for each of Days 1 and 2 are the sums of the joint probabilities over the columns and rows, respectively, of Table 4.5. Note that the probability of either Day 1 or Day 2 being sunny, cloudy, or rainy is 0.5, 0.3, or 0.2, respectively. This example illustrates the general result that a marginal probability is the *unconditional probability* of an event. As the term indicates, the unconditional probability of an event E is simply the probability of E , calculated without knowing whether other events related to E may or may not have occurred.

The probability, marginal probability, and unconditional probability of an event always have the same value.

4.7 Conditional probability**conditional probability**

Two events E_1 and E_2 are often related in such a way that the probability of occurrence of one depends on whether the other has occurred. The *conditional probability* of one event, given that the other has occurred, is equal to the joint probability of the two events divided by the marginal or unconditional probability of the given event. Thus, the conditional probability of event E_2 , given that event E_1 has occurred, denoted $Pr\{E_2|E_1\}$, is defined as:

$$Pr\{E_2 | E_1\} = \frac{Pr\{E_1 \cap E_2\}}{Pr\{E_1\}} \text{ for } Pr\{E_1\} > 0. \quad (4.14)$$

➤ If $Pr\{E_1\} = 0$, $Pr\{E_2 | E_1\}$ is undefined.

This formula for conditional probability can be applied in Example 4.9, which is an extension of Example 4.8.

Example 4.9. Weather on successive days. Suppose that the observed frequencies for 100 pairs of successive days are exactly equal to the theoretical frequencies implied by Table 4.5. For example, the number of two successive sunny days is exactly 40, because the probability of two successive sunny days is 0.40. Table 4.6 gives the frequencies of all nine weather combinations, along with the marginal frequencies.

Table 4.6. Joint and marginal frequencies for weather on 100 pairs of successive days

Day 1	Day 2			Marginal frequency (Day 1)
	Sunny (S_2)	Cloudy (C_2)	Rainy (R_2)	
Sunny (S_1)	40	5	5	50
Cloudy (C_1)	5	15	10	30
Rainy (R_1)	5	10	5	20
Marginal frequency (Day 2)	50	30	20	100

Note that Day 1 is sunny a total of 50 times. Of these 50 times, Day 2 is also sunny 40 times. Therefore, based on the frequencies from Table 4.6, the observed or empirical conditional probability of a sunny day given that the preceding day is sunny is $40/50 = 0.80$. Formally,

$$\begin{aligned} Pr\{S_2 | S_1\} &= (100)(0.40)/(100)(0.50) = 0.40/0.50 \\ &= Pr\{S_1 \cap S_2\} / Pr\{S_1\} \end{aligned} \quad (4.15)$$

Note that Equation (4.15) yields the same result as Equation (4.14), the defining equation for conditional probability.

Based on Table 4.5 and using Equation (4.15), Table 4.7 gives the conditional probabilities of the weather on Day 2 given the weather on Day 1.

Table 4.7. Conditional probabilities of weather on successive days

Day 1	Day 2		
	$Pr\{S_2 \mid \text{Day 1 weather}\}$	$Pr\{C_2 \mid \text{Day 1 weather}\}$	$Pr\{R_2 \mid \text{Day 1 weather}\}$
Sunny (S_1)	0.80	0.10	0.10
Cloudy (C_1)	0.17	0.50	0.33
Rainy (R_1)	0.25	0.50	0.25

By interchanging the roles of E_1 and E_2 in Equation (4.13), the joint probability of E_1 and E_2 can be written in two ways in terms of conditional probabilities:

$$\begin{aligned} Pr\{E_1 \cap E_2\} &= Pr\{E_1 \mid E_2\}Pr\{E_2\} \\ &= Pr\{E_2 \mid E_1\}Pr\{E_1\} \end{aligned} \quad (4.16)$$

Let E_1 and E_2 be any two events in a sample space S . From the properties of set operations in Table 4.2, E_2 can be written as

$$E_2 = E_2 \cap S = E_2 \cap (E_1 \cup E_1^c) = (E_2 \cap E_1) \cup (E_2 \cap E_1^c) \quad (4.17)$$

Because the two events in parentheses on the right side of Equation (4.17) are disjoint, it follows from Equation (4.2) that:

$$Pr\{E_2\} = Pr\{E_2 \cap E_1\} + Pr\{E_2 \cap E_1^c\} \quad (4.18)$$

Using Equation (4.15), Equation (4.18) can be written as:

$$Pr\{E_2\} = Pr\{E_2 \mid E_1\}Pr\{E_1\} + Pr\{E_2 \mid E_1^c\}Pr\{E_1^c\} \quad (4.19)$$

law of total probability for any events E_1 and E_2 . This formula is called the **law of total probability**. It is the basis for quantifying event trees, which are frequently used to diagram the possibilities in an accident sequence.

From Equation (4.13), the joint probability of two statistically independent events is the product of their individual unconditional probabilities. If E_1 and E_2 are statistically independent, it follows from Equation (4.16) that, if neither E_1 nor E_2 has probability 0, then equations (4.20) and (4.21) hold.

$$Pr\{E_1\} = Pr\{E_1 | E_2\} \quad (4.20)$$

$$Pr\{E_2\} = Pr\{E_2 | E_1\} \quad (4.21)$$

From these equations, if two events are statistically independent, the probability of either event does not depend on whether the other event occurs. If two events are not statistically independent, they are said to be **statistically dependent events**.

The concepts of mutually exclusive events and statistically independent events are often confused. If E_1 and E_2 are mutually exclusive events, $Pr\{E_1 \cap E_2\} = Pr\{\emptyset\} = 0$. If neither event has probability 0, from Equation (4.13), $Pr\{E_1 \cap E_2\} \neq 0$. Thus, the two events are not independent. Mutually exclusive events cannot be independent and vice versa. Another explanation is that mutually exclusive events cannot be independent because the occurrence of either one precludes the occurrence of the other.

Equation (4.16) can be generalized to calculate the probability of the intersection of any number of events (the probability that all the events occur simultaneously). In general, the probability of the simultaneous occurrence of n events can be written as:

$$Pr\{E_1 \cap E_2 \cap \dots \cap E_n\} = Pr\{E_1\}Pr\{E_2 | E_1\}Pr\{E_3 | E_2 \cap E_1\} \dots Pr\{E_n | E_{n-1} \cap \dots \cap E_1\} \quad (4.22)$$

Equation (4.22) is referred to as the **chain rule of probability**. This rule can be used to calculate the probability that a given accident sequence occurs, with E_1 denoting the initiating event and the remaining events corresponding to the failures or successes of the systems that must function in order to mitigate such an accident.

If E_1, E_2, \dots, E_n are statistically independent events, the probability that at least one of the n events occurs is equal to 1 minus the probability that none of the n events occurs. Formally,

$$Pr\{E_1 \cup E_2 \cup \dots \cup E_n\} = 1 - (1 - Pr\{E_1\})(1 - Pr\{E_2\}) \dots (1 - Pr\{E_n\}) \quad (4.23)$$

Expansion of the right side of Equation (4.23) leads to the general inclusion-exclusion principle, as illustrated in Equation (4.11) for $n = 3$. For the simple case where $Pr\{E_1\} = Pr\{E_2\} = \dots = Pr\{E_n\} = p$, the right side of Equation (4.23) reduces to $1 - (1 - p)^n$.

Probabalistic Risk Assessment The general result in Equation (4.22) has applications in fault tree analysis used in **Probabalistic Risk Assessment** (PRA). For example, for a system where system failure occurs if any one of n independent events occurs, Equation (4.23) gives the probability of system failure. One such example could be failures of critical system components. In general, the events represent the modes by which system failure (the top event of the fault tree) can occur.

minimal cut set These modes are referred to as the **minimal cut sets** of the fault tree. The events are independent if no minimal cut sets have common component failures. See Vesely et al. (1981) for further discussion of fault trees and minimal cut sets.

If the n events are not independent, the right side of Equation (4.23) may be larger than or smaller than the left side. However, for an important situation that frequently arises in PRA, the right side of Equation (4.23) forms an upper bound for the left side. If the n events are cut sets that are positively associated, then the right side is an upper bound for $Pr\{E_1 \cup E_2 \cup \dots \cup E_n\}$ and is known as the **min cut upper bound** (NRC (2003), p. A-3). This designation arises from common PRA applications where E_i is the i^{th} minimal cut set of a system or accident sequence of interest. In this case, the min cut upper bound is superior to the rare event approximation in Equation (4.12) and can never exceed unity (as can happen with the rare event approximation). If the n events satisfy conditions similar to those of the rare event approximation, the min cut upper bound is a useful approximation. Note that, because they are negatively associated, the min cut upper bound is not applicable to mutually exclusive events.

4.8 Bayes' Theorem

In PRA applications, the true values of many input parameters are unknown and must be estimated. If relevant data are available, either from observation or experimentation, a common approach is to calculate the probabilities of the possible input parameters given the observed data. In a probability framework, this means calculating the conditional probability of an event A given that another event B has occurred. In the PRA context, A represents a specified hypothesis or state of nature and B is the observed data. This calculation is carried out using **Bayes' Theorem**, a general result in the theory of conditional probability.

**Bayes
Theorem**

Bayes' Theorem states that if $\{A_1, A_2, \dots, A_n\}$ is a partition of the sample space (see Equation (4.4)) and if B is any event such that $Pr\{B\} > 0$, then, for all $i = 1, 2, \dots, n$,

$$Pr\{A_i | B\} = \frac{Pr\{B | A_i\}Pr\{A_i\}}{Pr\{B\}} \quad (4.24)$$

where the denominator of Equation (4.24) is given by:

$$Pr\{B\} = \sum_{j=1}^n Pr\{B | A_j\}Pr\{A_j\} \quad (4.25)$$

Equation (4.24) follows from the relations in Equation (4.14), which can be written as:

$$Pr\{A_i | B\}Pr\{B\} = Pr\{B | A_i\}Pr\{A_i\} \quad (4.26)$$

Equation (4.22) follows by using Equation (4.5) with $E = B$ and $S_j = A_j$ and then using Equation (4.16) with $E_1 = B$ and E_2 replaced by A_j .

<i>prior probability</i>	In Bayes' Theorem, the partition $\{A_1, A_2, \dots, A_n\}$ represents all possible states of nature or hypotheses and B represents data from observation or experimentation. $Pr\{A_i\}$ is the prior probability of A_i before the data B are known. The conditional probability $Pr\{B A_i\}$ is the probability of observing B if A_i is true.
<i>posterior probability</i>	The conditional probability $Pr\{A_i B\}$ is the posterior probability of A_i once B is known. The denominator $Pr\{B\}$ of Equation (4.24) is the unconditional probability of the data averaged over all possible hypotheses or states of nature and serves as a normalizing constant.
<i>Bayesian inference</i>	As Equation (4.24) indicates, the posterior probability $Pr\{A_i B\}$ for a specified value of i depends not only on the prior probability of A_i and the conditional probability of the data given A_i , but also on all the other prior probabilities and conditional probabilities of the data. Thus, Bayes' theorem is the basis for Bayesian inference , which refers to the process of converting the prior distribution of the partition $\{A_1, A_2, \dots, A_n\}$ to a posterior distribution.

More on Bayesian inference may be found in Chapter 20.

4.9 Probability Estimation

This chapter has focused on the basic definitions and relationships among the probabilities of various types of events. The numerical examples presented have all been based on assumed values for the probabilities involved. In practice, however, probability estimation is usually based on information from observation, experimentation, or judgment. The assignment of numerical values to event probabilities is carried out by a combination of one or more of three approaches: modeling, data-based probability, or subjective probability.

It must be emphasized that all of the approaches outlined below have associated uncertainty. Uncertainty is associated with the assumptions required for each approach, and statistical uncertainty is associated with the data. The characterization and evaluation of statistical uncertainty are the subjects of much of this book. However, while no less important, the uncertainty associated with the assumptions is beyond the scope of this book. Nevertheless, whatever the approach or combination of approaches used for probability estimation in any application, a complete analysis requires an assessment of all the uncertainties.

modeling approach to probability

The simplest **modeling approach to probability** is to assume a probability model that allows the probability of any event to be calculated using the axioms of probability without any reference to data or experimentation. The simplest assumption is that all outcomes or points in the sample space are equally likely. If there are N points in the sample space, then each point must then have probability $1/N$. It follows immediately that, if an event E consists of k outcomes, $Pr\{E\} = k/N$. Examples are tossing any number of fair coins or fair dice.

Another modeling approach is to assume a probability model with a completely specified structure, which allows the calculation of the probabilities using the axioms of probability. Examples are radioactive decay of a known isotope or reliability of a component with an extensive failure history.

Although assuming a completely specified probability model allows the exact calculation of the probability of any event, this approach has limited applicability to practical problems. A more common modeling approach is to assume a probability model with a specific structure but with one or more unknown parameters. Examples are a normal model (Chapter 6) or a Poisson model (Chapter 23) with unknown means. With this approach, it is necessary to estimate the unknown parameter(s), but once this is done, the estimated probability of any event can be calculated.

data-based approach to probability

The **data-based approach to probability** uses observed or experimental data together with statistical estimation techniques (Chapter 9) to estimate unknown probabilities or parameters. For example, if an experiment is independently repeated n times and the event E occurs n_E times, then an estimate of $Pr\{E\}$ is n_E/n .

relative frequency

This is the **relative frequency** estimate of $Pr\{E\}$. More generally, data are used to estimate the parameters of assumed probability models (e.g., the mean of an assumed binomial distribution used to model component reliability). Because such estimates depend on the amount of available data as well as their random characteristics, uncertainty is associated with them.

It is important to clarify a common misconception about the relative frequency estimate of $Pr\{E\}$. The limiting value of n_E/n as $n \rightarrow \infty$ is sometimes taken as the definition of $Pr\{E\}$. This approach has two problems. First, to define a probability as a limiting value is highly problematic because it cannot be evaluated in the real world, and furthermore, it is necessary to assume that such a limit exists. Second, this definition is not a proper definition because it is a theorem in the theory of probability (i.e., it is a consequence of the axioms of probability).

subjective probability approach Another approach to estimating probabilities is the **subjective probability approach** (see Chapter 20). Probabilities obtained from the opinions of people are examples of subjective probabilities. Here, probability can be thought of as a measure of belief. Any past information about the problem being considered can be used to help estimate the various probabilities. In particular, information about the relative frequency of occurrence of an event could influence the results.

proposition Martz and Waller (1991, Chapter 5) present subjective probability as dealing not only with repeatable events but also with **propositions**. A proposition is considered to be a collection of events that cannot be conceived as a series of repetitions (e.g., a meltdown at a specified nuclear power plant). The degree of belief in proposition A , the subjective probability $Pr\{A\}$, represents how strongly A is believed to be true. At the extremes, if A is believed to be certain or true, $Pr\{A\} = 1$; if A is believed to be impossible or false, $Pr\{A\} = 0$. Values between 0 and 1 represent intermediate beliefs between false and true.

The subjective probability approach is often used in Bayesian inference to estimate the prior probabilities used to calculate the posterior probabilities (Section 4.8). Accordingly, the posterior probabilities must also be considered as subjective probabilities. While Bayesian inference can certainly be carried out without using subjectively determined priors (e.g., differential diagnosis of possible diseases based on a large medical database of symptoms and test results), for many PRA applications, it is necessary to estimate the prior probabilities using the subjective probability approach.

There are two sources of uncertainty in the subjective probability approach. The first stems from uncertainty about the chance of occurrence of an event in the real world based on someone's subjective judgment. The second stems from the diversity of results when more than one subjective probability is elicited (e.g., when a panel is used in an expert elicitation process). Consideration of these uncertainties is beyond the scope of this book.

5

Errors

5.1 What to look for in Chapter 5

Chapter 5 alerts us to the sources and characterization of errors we may encounter in everyday life and in data. These underlie the statistical methods presented in future chapters. The concepts discussed in this chapter are:

- *measurement system*, §5.3
- *metrology*, §5.3
- *error*, §5.3
- *accuracy*, §5.3
- *bias, systematic error*, §5.3
- *precision*, §5.3
- *uncertainty*, §5.4

5.2 Errors, errors, ... everywhere

Did this ever happen to you?

You weigh yourself in the morning. You don't like the reading of the scale. You rationalize that the scale is in error. Further, you decide that this dubious

measurement could be the result of one or more of the following errors⁷: instrument error, resolution error, calibration error, replication error, sampling error, accuracy error, design error, day-of-the-week error, time-of-the-day error, seasonal variation error, rounding-off error, or “something I ate” error.

Some of these listed errors are clearly redundant, some are clearly irrelevant, and some possible errors are missing from the list. As a rule, we consider all the errors we can imagine that are relevant and then weed down the list. For example, we calculate the air mass inside a reactor containment from measured physical variables, such as pressure, temperature, and relative humidity. However, we should consider whether we have left out any other variables, such as outdoor temperature changes, voltage fluctuations, and neighborhood air traffic, just in case any of those affects the air mass measurements. Then, using engineering and statistical judgment, we focus on those effects that are relevant. In this way, we may reduce the error that would otherwise be ignored.

If we read one number on a scale but report another, we made a mistake. Mistakes can be accidental or intentional (e.g., a deliberately erroneous report intending to deceive). If we calculate the containment air mass and change the record, part of the record, or the final calculation, we have made a mistake. This book does not deal with mistakes, although some mistakes can be detected statistically or mathematically. We assume the database given to us to analyze is free of mistakes. We conduct an honest analysis based on what we believe is an honest collection of information.

5.3 Characterizing errors: accuracy and precision

measurement *Measurement* is the assignment of a number to some characteristic or property of an object such as its length, weight, or temperature. Measurements are typically obtained by employing a measuring device such as a yardstick, scale, or thermometer. Measurements always relate to properties of objects, not to the objects themselves. Hence, we don’t measure a fuel rod, but we do measure its length, weight, and enrichment.

metrology The science that studies and codifies the measurement process is called *metrology*. Metrology is concerned not only with the measurement of a property but also with the error associated with such a measurement and the probability that the measurement is within a predetermined distance from the true value of the property. Chapter 9, “Statistical Estimation,” presents a detailed discussion of error quantification.

⁷ The following section presents a formal definition of error. For now, an intuitive understanding of error will suffice.

The measurement device is not the only source of error in a study or experiment. In addition to the value of the characteristic we wish to measure, the measurement, and consequently the error depend on a combination of other factors. These factors include errors stemming from the physical characteristics of a sensor, its fabrication, the technician who reads the sensor, and the clerk who writes down and transmits the reading. All of the factors that contribute to a measurement and its associated error are called a **measurement system**.

A measurement is designed to measure the **true value** of the property being measured. Because of measurement error, the true value of a property may never be exactly determined. However, it is conceptually useful in modeling a measurement system. The concept of a true value is analogous to the concept of a probability as discussed in Section 4.2. Although true values and probabilities associated with physical objects can never be determined, they are both conceptually useful for modeling.

In this discussion, a true value is designated by the Greek letter τ (tau). Let Y be a measurement produced by a measurement system. That measurement may or may not be “close” to τ . The difference between Y and τ is called an **error**, which is traditionally denoted by the Greek letter ε (epsilon). We write that difference as:

$$\varepsilon = Y - \tau \quad (5.1)$$

Consider a sequence of measurements $\{Y_1, Y_2, \dots, Y_i, \dots\}$ produced by a measurement system and their corresponding errors $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i, \dots\}$. The concepts of **accuracy** and **precision** are defined in terms of the behavior of the errors in the sequence $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i, \dots\}$.

Each measurement system has an associated accuracy. Accuracy is assessed in terms of the mean of ε . Because ε is a random variable (to be discussed in Chapter 6), its mean has not yet been defined. However, for the purpose of this discussion, the mean of ε can be thought of as a long-run average (i.e., the arithmetic mean of a large number of errors (see Section 2.3)). If the mean of ε is zero, then the system is said to be accurate. If the mean is not zero, the system is said to be inaccurate.

The accuracies of two or more measurement systems are often compared. Consider a second measurement system, also aimed at determining the same true value, τ . If the mean of the error of the first system is closer to zero than the

mean of the error of the second system, then the first system is said to be more accurate than the second system.

The concept of accuracy can be illustrated by rifle target practice. Even if we aim carefully at the target, sometimes we shoot to the left of the target and sometimes to its right, sometimes high and sometimes low. If, on the average, we score a bull's-eye, the system (we, the rifle, the weather, the shooting range, etc.) is considered accurate. Note that the measurement system can be accurate, even if we never hit the bull's-eye.

bias A measurement system is said to have a **bias** (or to be biased) if it is inaccurate. Whether the magnitude of the bias is critical depends on the system and the use to which it is applied. If the long-run average of the measurements is τ' , the difference $\tau' - \tau$ is the system bias. Mathematically, the bias may be written as:

$$b = \tau' - \tau \quad (5.2)$$

unbiased If $b = 0$, the measurement system is said to be **unbiased**. Equivalently, the system is unbiased if the long-run average of the errors is zero (i.e., the system is accurate).

systematic error Bias is sometimes referred to as **systematic error**.

precision A second characteristic of a measurement system is **precision**. The idea behind precision is how repeatable the measurements are (i.e., how close the measurements are to each other). The precision of a series of measurements may be measured by the variance of Y . Because Y , like ε , is a random variable (to be discussed in Chapter 6), its variance has not yet been defined. However, for the purpose of this discussion, the variance of Y may be thought of as the long-run sample variance (i.e., the sample variance of a large number of measurements (see Section 2.8)). In this sense, one measurement system is said to be more precise than a second system if it has a smaller variance than the second system.

As with accuracy, the concept of precision may be illustrated by rifle target practice. Regardless of where our bullets hit, the system is precise if all the hits are in exactly the same place. Clearly, if we and our rifle are simultaneously accurate and precise, we have a statistically superb system.

Figure 5.1 displays four combinations of good/poor accuracy and good/poor precision in a one-dimensional measurement system.

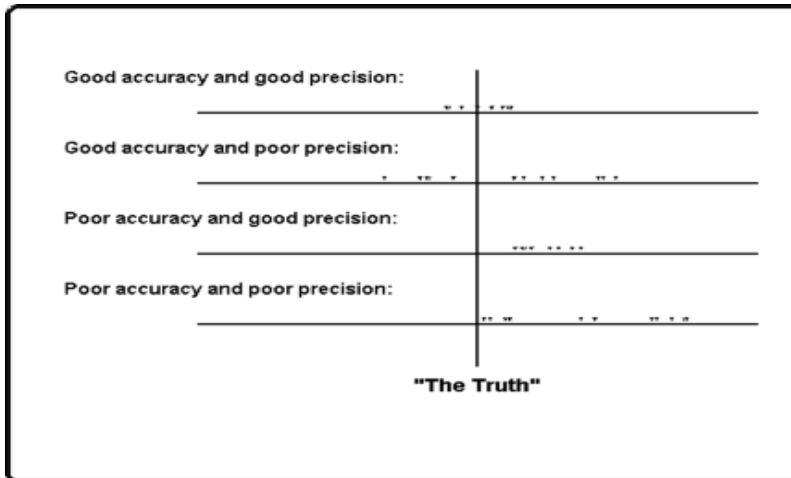


Figure 5.1. Accuracy and precision illustrated

The combinations of accuracy and precision illustrated in Figure 5.1 do not tell the whole story. In particular, suppose we have an accurate and precise measurement system with a great deal of variability in the error about the true value. This would be illustrated by the top line in Figure 5.1 with the measurements much more spread out. In such a case, we would not have a “good” measurement system.

5.4 Uncertainty

uncertainty Related to accuracy (or the lack of) and precision (or the lack of) is the concept of *uncertainty*. Uncertainty is broadly used as a measure of the “incorrectness” in a set of data arising from inaccuracy, imprecision, variability, or some combination of these. The term itself is troublesome because, unless it is precisely defined, it may convey a broad spectrum of meaning and suffer from an equally broad spectrum of interpretation. The term itself does not have an accepted meaning in statistics. One global measure of uncertainty that is sometimes used is that of mean-square error (see Section 6.10). On the whole, it seems best to avoid using “uncertainty” in technical discussions unless it is carefully defined and, if possible, quantified.

6

Random variables

6.1 What to look for in Chapter 6

A random variable is a quantification of a probability model that allows us to model random data. It is an essential element in the application of statistical methods to data. The concepts discussed in this chapter are:

- *random variable discrete and continuous*, § 6.2
- *probability mass function (pmf)*, § 6.3
- *probability function (pf)*, § 6.3
- *cumulative distribution function (cdf)*, § 6.3
- *probability density function (pdf)*, § 6.4
- *bivariate distribution, joint distribution*, § 6.5, § 6.6
- *marginal distribution*, § 6.5
- *conditional distribution*, § 6.5
- *probabilities of independent events*, § 6.6
- *expected value, mathematical expectation*, § 6.7
- *linear combinations and their variance*, § 6.8
- *standard error of the mean*, § 6.8
- *contrast and linear contrast*, § 6.9

6.2 Random variables

An outcome of an experiment may be either a quantitative or a qualitative variable. Because statistical methods deal with numerical data, it is necessary to assign numerical values to the outcomes of the experiment in order to analyze the results. A **random variable** is a function that assigns a unique numerical value to each outcome of an experiment along with an associated probability. We distinguish between two types of random variables: discrete and continuous. A **discrete random variable** corresponds to an experiment that has a finite or countably infinite (e.g., the set of positive integers) number of outcomes. (Continuous random variables will be discussed following the examples presented below.)

In terms of the concepts introduced in Chapter 4, a discrete random variable is a quantification of a special case of a probability model defined in Section 4.5. Let $S = \{E_1, E_2, \dots, E_i, \dots\}$ be a sample space such that all the sample points are disjoint elementary events. Let $p_i = Pr\{E_i\}$, where $\sum p_i = 1$. To quantify S , let $Y = \{y_1, y_2, \dots, y_i, \dots\}$ be a function of S (i.e., y_i is a real number corresponding to the point E_i). Then Y is a discrete random variable with probability distribution $\{p_i\}$.

By convention, we denote random variables by capital letters and the values they take on by lower case letters. Thus, the diameter of a fuel rod is denoted by the random variable Y , but once it is measured, the diameter is denoted by y , (e.g., $y = 1.07$ cm).

Example 6.1. Tossing a fair coin. The sample space consists of two points, T and H , each with probability $p = 1/2$. Define a random variable Y to be 0 if T occurs and 1 if H occurs. The random variable Y is the number of heads in one toss of a fair coin.

Example 6.2. Rolling a fair die. The sample space consists of the six points corresponding to the six faces of the die, each with probability $p = 1/6$. Define a random variable Y to be equal to the number of spots showing on the face of the die. The random variable Y takes on the values 1, 2, ..., 6.

Example 6.3. Valve status. The sample space consists of the two points corresponding to two positions of the valve: closed, with probability p , and open with probability $1 - p$. Define a random variable Y to be 0 if the valve is closed and 1 if it is open.

continuous random variable A **continuous random variable** is a random variable for which the set of possible outcomes (i.e., the sample space) is continuous. Examples of continuous sample spaces are weight, length, and time to failure; these may assume any value in the interval between 0 and ∞ . A continuous random variable may also be negative, (e.g., temperature, loss in weight, or a loss in the value of a share in the stock market). A formal mathematical definition of a continuous random variable is beyond the scope of this book.

In the strictest sense, measurements are never continuous; they are always discrete. For example, time is measured to the nearest hour, minute, second, or fraction of a second, depending on the application. Therefore, all random variables are necessarily discrete. However, because it is generally easier to deal mathematically with continuous rather than discrete random variables, discrete random variables are often replaced by continuous random variables for modeling and analysis purposes. For example, if the time to failure of a pump is measured to the nearest hour, then time to failure is a discrete random variable taking on the values 1, 2, However, for analysis purposes, it may be convenient to assume that time to failure is a continuous random variable taking on any positive value. Most importantly, in replacing a discrete random variable with a continuous random variable, we must be sure that the effect of the replacement on the result of the analysis is negligible.

6.3 Distributions of discrete random variables

probability mass function For a discrete random variable Y , the **probability mass function (pmf)** gives the probability associated with each possible value y of Y . That probability may be denoted by $Pr\{Y = y\}$, $Pr\{y\}$, or $f(y)$. Consistent with the definition of probability, we have:

$$Pr\{y\} \geq 0 \quad (6.1)$$

$$\sum Pr\{y\} = 1 \quad (6.2)$$

probability function, distribution When there is no ambiguity, a probability mass function is sometimes referred to as a **probability function** or a **distribution**.

Two examples of discrete random variables are given below. Note that all probabilities are nonnegative and sum to 1, in accordance with Equations (6.1) and (6.2).

Example 6.4. Sum of dots. Suppose two fair dice are rolled. Let Y be the total number of dots. Table 6.1 gives the values, y , of Y and their associated probabilities, $Pr\{y\}$. (The second and third rows of Table 6.1 will be considered later in this section.) Table 6.1 also highlights three selected values of y ($y = 8$, $y = 9$, and $y = 10$) and their probabilities. If in this example we wish to find the

probability of obtaining any of these (mutually exclusive) values, we add their probabilities to get $Pr\{8 \leq y \leq 10\} = 5/36 + 4/36 + 3/36 = 1/3$.

Table 6.1. Probability distribution for the sum of two fair dice

Y	1	2	3	4	5	6	7	8	9	10	11	12
$Pr\{Y = y\}$	0	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
$F(y)$	0	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	1
$1 - F(Y)$	1	35/36	33/36	30/36	26/36	21/36	15/36	10/36	6/36	3/36	1/36	0

The probabilities in Table 6.1 are plotted in a bar chart in Figure 6.1, where the three shaded columns correspond to the selected probabilities in Example 6.4.

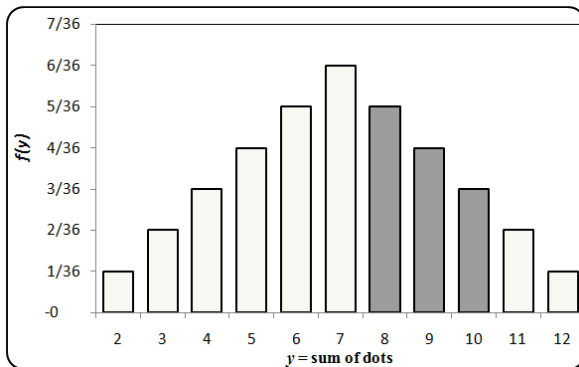


Figure 6.1. Distribution of sum of dots for two fair dice

Example 6.5. Snubber failure. Let Y be the number of snubbers in a population of 1000 snubbers that fail during the first 5 years of operation. Table 6.2 gives a hypothetical probability distribution for Y .

To find the probability that one or more snubbers will fail in the first 5 years, we calculate $Pr\{1\} + Pr\{2\} + Pr\{3\} = 0.10 + 0.03 + 0.02 = 0.15$. Alternatively, since the probabilities sum to 1, we calculate this probability as 1 less than the probability of 0 failures, or $1 - Pr\{0\} = 1 - 0.85 = 0.15$.

Table 6.2. Probability distribution for snubber failures

Y	0	1	2	3
$Pr\{Y = y\}$	0.85	0.10	0.03	0.02
$F(y)$	0.85	0.95	0.98	1.00
$1 - F(y)$	0.15	0.05	0.02	0.00

The probabilities in Table 6.2 are plotted in Figure 6.2, where the shaded columns correspond to the probabilities selected in the example.

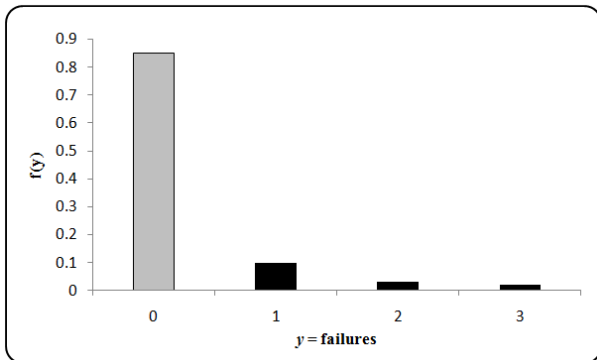


Figure 6.2. Distribution of snubber failures

Note that a function of a discrete random variable Y is also a discrete random variable, with a distribution determined by the distribution of Y . For instance, in Example 6.4, let $Z = Y^2$. That is, Z is the square of the sum of dots. Then Z takes on the 11 values 4, 9, ..., 144, with distribution given by the second row in Table 6.1, labeled $Pr\{Y = y\}$.

cumulative distribution function (cdf)

Probability functions provide point probabilities for discrete random variables. A related function is the **cumulative distribution function (cdf)**, denoted by $F(y)$ and defined as the probability that the random variable takes on values less than or equal to the value y . For a discrete random variable Y , $F(y)$ is the sum of the probabilities of all values of $Y \leq y$. Mathematically,

$$F(y) = Pr\{Y \leq y\} = \sum_{Y \leq y} Pr\{y\} \tag{6.3}$$

The *cdf* for Example 6.4 is shown in Table 6.1 in the row labeled $F(y)$.

Table 6.2 shows the $F(y)$ values for Example 6.5. From that table, we see that the probability of 0 or 1 snubber failure in 5 years is $F(1) = 0.95$.

- When there is no ambiguity, the *cdf* is also called the distribution. Thus, the term distribution applies to both the *pmf* and the *cdf*.

The *cdf* for Example 6.4 is plotted from the line labeled $F(y)$ in Table 6.1. In this case, Figure 6-3 may be easier to read and interpret than the table. For example, the probability of the sum of the dots in a throw of two fair dice being as much as 4 is $F(4) = 6/36 = 1/6$.

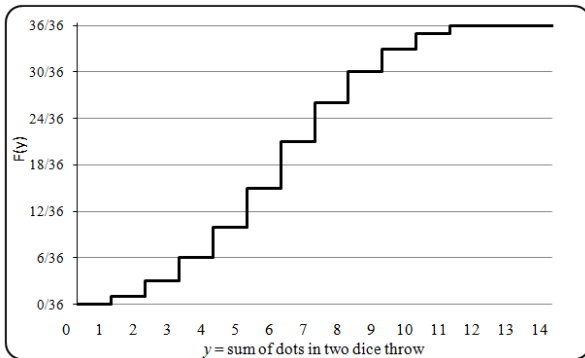


Figure 6.3. Cumulative distribution function for Example 6.4

In Figure 6-4, the *cdf* for Example 6.5 is plotted from the line labeled $F(y)$ in Table 6.2. For example, the probability of two or fewer snubber failures is 0.98.

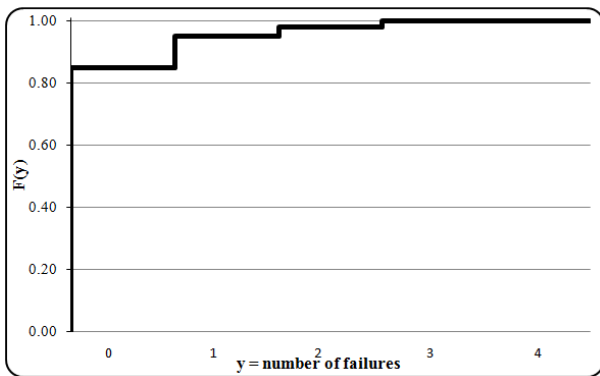


Figure 6.4. Cumulative distribution function for Example 6.5

complementary cumulative distribution function A sometimes useful function related to the *cdf* is the **complementary cumulative distribution function**. For a value y , this function is defined as $1 - F(y)$. From Equation 6.3, $F(y) = Pr\{Y > y\}$. In Example 6.4 and Table 6.1, the values of

$1 - F(y)$ are shown in the last line of Tables 6.1 and 6.2, respectively.

Figure 6.5 shows the complementary *cdf* for Example 6.4. Here we see that the probability of two or more snubber failures is only 0.02, or 2%.

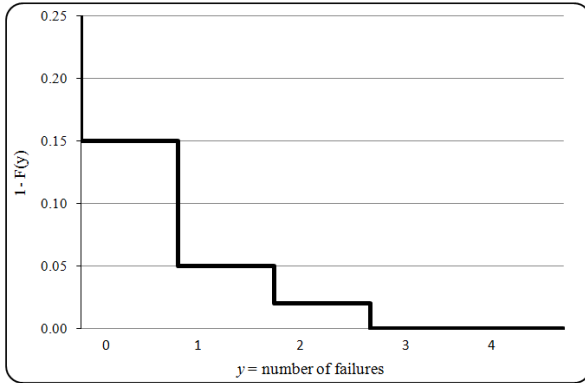


Figure 6.5. Complementary cumulative distribution function for Example 6.5

6.4 Distributions of continuous random variables

probability density function density

Whereas a discrete random variable can take on only some of the values in an interval, a continuous random variable Y may take on any value in an interval. In parallel with the discrete mass function, we have a **probability density function (pdf)**, or simply **density**, also denoted by $f(y)$. The density is defined as a nonnegative integrable function with the following properties:

$$f(y) \geq 0 \tag{6.4}$$

$$\int_{-\infty}^{+\infty} f(y)dy = 1 \tag{6.5}$$

For any continuous random variable Y and for any two points a and b where $a < b$, the density has the following property:

$$\int_a^b f(y)dy = Pr\{a \leq y \leq b\} \tag{6.6}$$

This equation is illustrated in Figure 6.6. As Equation (6.6) indicates, the shaded area is the probability that a continuous random variable Y falls between the values of a and b .

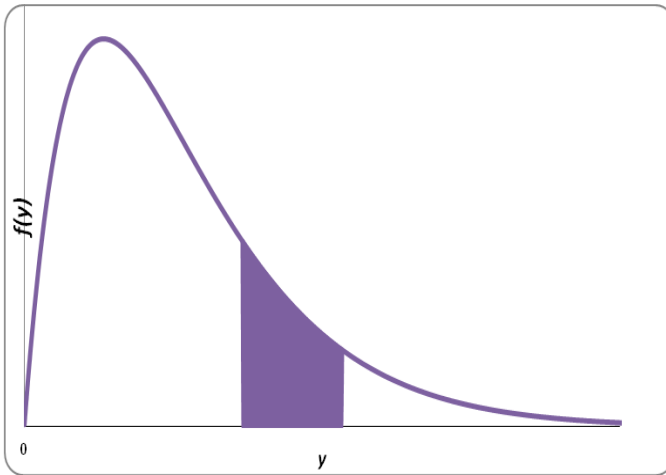


Figure 6.6. Probability of occurrence in interval (a, b)

In parallel with the probability mass function, continuous random variables also have a cumulative distribution function (*cdf*). This function, also denoted $F(y)$, is defined as the probability that the continuous random variable takes on values less than (or equal to) the value y . Mathematically,

$$F(y) = Pr\{Y \leq y\} = \int_{-\infty}^y f(y)dy \quad (6.7)$$

Here are some relevant points:

- Since $F(y)$ represents probability, $0 \leq F(y) \leq 1$.
- $F(y)$ is a nondecreasing function of y , that is, if $y_1 \leq y_2$, then $F(y_1) \leq F(y_2)$.
- $f(y)$ is the first derivative of $F(y)$ with respect to y .
- If Y takes on only positive values, the limits of integration are 0 to y .
- If Y has a restricted range, with $a \leq Y \leq b$, then $F(a) = 0$ and $F(b) = 1$.
- For both discrete and continuous random variables:

$$Pr\{y_1 < Y \leq y_2\} = F(y_2) - F(y_1) \quad (6.8)$$

If Y is a continuous random variable, Equation (6.8) implies that:

$$Pr\{Y = y\} = 0 \text{ for any } y \quad (6.9)$$

We are often concerned with extreme outcomes, as they may present an unusual risk or a specific interest. For concerns about the probability of obtaining a small value of Y (say smaller than or equal to a), we investigate $Pr\{Y \leq a\}$, which is $F(a)$. For concerns about the probability of obtaining a large value of Y (say larger than or equal to b), we investigate $Pr\{b \leq Y\}$, which is $1 - F(b)$.

Parallel to discrete random variables, the term $1 - F(y)$ is the complementary distribution function of Y .

6.5 Distributions of discrete bivariate random variables

**joint probability
function**

The concepts of joint, marginal, and conditional probability, discussed in Chapter 4, also pertain to random variables and probability distributions. To apply these concepts we first consider two discrete random variables, X and Y , where X may take on any of m values x_1, x_2, \dots, x_m , and Y may take on any of n values y_1, y_2, \dots, y_n . The probability that X will take on a value x_i , and, simultaneously, Y will take on a value y_j , is given by the **joint probability function** (joint *pf*), denoted by $f(x_i, y_j)$.

$$f(x_i, y_j) = Pr\{X = x_i, Y = y_j\} \quad (6.10)$$

where

$$f(x_i, y_j) \geq 0 \quad (6.11)$$

$$\sum_i \sum_j f(x_i, y_j) = \sum_{i,j} f(x_i, y_j) = 1 \quad (6.12)$$

**discrete bivariate
random variable** The ordered pair (X, Y) is a **discrete bivariate random variable** with a joint *pf* given by Equation (6.10).

The joint *pf* is tabulated in the shaded area of Table 6.3.

Table 6.3. Joint and marginal distributions of (X, Y)

	$Y = y_1$	$Y = y_2$	\dots	$Y = y_n$	Marginal
$X = x_1$	$f(x_1, y_1)$	$f(x_1, y_2)$	\dots	$f(x_1, y_n)$	$f_X(x_1)$
$X = x_2$	$f(x_2, y_1)$	$f(x_2, y_2)$	\dots	$f(x_2, y_n)$	$f_X(x_2)$
	\vdots	\vdots	\vdots	\vdots	\vdots
$X = x_m$	$f(x_m, y_1)$	$f(x_m, y_2)$	\dots	$f(x_m, y_n)$	$f_X(x_m)$
Marginal	$f_Y(y_1)$	$f_Y(y_2)$	\dots	$f_Y(y_n)$	1

**(marginal)
probability
function**

The probability that X will take on a specific value (say x_i) is calculated as the sum all the probabilities $f(x_i, y_j)$ that involve x_i , regardless of the associated values y_j of Y . This sum is called the **(marginal) probability function** of X evaluated at $X = x_i$ and is denoted by $f_X(x_i)$. Mathematically,

$$f_X(x_i) = Pr\{X = x_i\} = \sum_j f(x_i, y_j) \quad (6.13)$$

**(marginal)
distribution**

The set $\{f_X(x_i)\}$ of marginal probabilities in the last column of Table 6.3 is the **(marginal) distribution** of X .

Similarly, the marginal probability function $f_Y(y_j)$ of Y is given by:

$$f_Y(y_j) = Pr\{Y = y_j\} = \sum_i f(x_i, y_j) \quad (6.14)$$

The set $\{f_Y(y_j)\}$ of marginal probabilities in the bottom row of Table 6.3 is the (marginal) distribution of Y .

As in Section 6.3, the discrete bivariate random variable (X, Y) has a cumulative distribution function, $F_{XY}(x, y)$. Here, $F_{XY}(x, y)$ is the probability that $X \leq x$ and $Y \leq y$. Symbolically,

$$F_{XY}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j) \quad (6.15)$$

The following example illustrates the joint and marginal distributions for discrete bivariate random variables.

Example 6.6. Sex discrimination. A large organization’s hiring practice is investigated to see whether there is statistical evidence of sex discrimination. Based on data collected over 5 years, the joint and marginal probabilities of hiring decisions for male and female applicants are shown in Table 6.4. Chapter 12 gives an analysis of count data that is related to this example. This example illustrates some of the concepts covered in this and subsequent chapters.

Table 6.4. Data for Example 6.6

	Gender		
Hiring status	Male (y = 1)	Female (y = 2)	Marginal probability
Not hired (x = 1)	0.24	0.36	0.60
Temporary (x = 2)	0.12	0.18	0.30
Permanent (x = 3)	0.04	0.06	0.10
Marginal probability	0.40	0.60	1.00

The joint probabilities $f(x,y)$ are shown in the shaded area, and the marginal probabilities $f_X(x)$ and $f_Y(y)$ are shown in the margins of the table. For example, the joint probability $f(3, 1)$ that an applicant will be both male and hired for a permanent position is 0.04. The marginal probability $f_X(3)$ of an applicant being hired for a permanent position is 0.10, and the marginal probability $f_Y(1)$ of an applicant being male is 0.40. In other words, there is a 10% chance that an applicant will be hired for a permanent position and a 40% chance that an applicant is male.

The concepts of joint and marginal distributions for discrete random variables can be extended to more than two discrete random variables. See Bowen and Bennett (1988), p. 76, for details

conditional probability

At times we may wish to know the probability that a discrete random variable Y will attain a specific value y_j if a second discrete random variable is known to have attained a value x_i . For example, we may wish to know the probability that an applicant will be hired for a permanent position if the applicant is known to be male. In the notation of Table 6.3, this is the **conditional probability** that $Y = y_j$, given that $X = x_i$. This type of probability, previously discussed in Section 4.7, is denoted by $g(y_j | x_i)$, is read as the “conditional probability of y_j given x_i ,” and is calculated by:

$$g(y_j | x_i) = \frac{f(x_i, y_j)}{f_X(x_i)} \quad (6.16)$$

The conditional probability in Equation (6.16) is defined only if $f_X(x_i) \neq 0$.

To find the probability that a male applicant in Example 6.6 will be hired for a permanent position, we calculate from Equation (6.16) and Table 6.4 that:

$$g(3|1) = \frac{f(3,1)}{f_Y(1)} = \frac{0.04}{0.40} = 0.10$$

That is, there is a 10% chance that a male applicant will be hired for a permanent position.

In analogy to Equation (6.16), the conditional probability of x_i given y_j , we have:

$$h(x_i | y_j) = \frac{f(x_i, y_j)}{f_Y(y_j)}, \quad f_Y(y_j) \neq 0 \quad (6.17)$$

conditional distribution The set of conditional probabilities $\{g(y_j | x_i)\}$ for all values of i and j is called the **conditional distribution** of Y given X . Similarly, the set of conditional probabilities $\{h(x_i | y_j)\}$ for all values of i and j is the conditional distribution of X given Y .

The definitions and calculations of conditional distributions may also be extended to more than two discrete random variables. See Bowen and Bennett (1988), p. 77, for details.

independent random variables Two discrete random variables X and Y are **independent random variables** if their joint distribution is equal to the product of the two marginal distributions. That is,

$$f(x_i, y_j) = f_X(x_i) f_Y(y_j) \quad (6.18)$$

The cumulative distributions of independent discrete random variables also multiply. That is,

$$F(x_i, y_j) = F_X(x_i) F_Y(y_j) \quad (6.19)$$

Equations (6.18) and (6.19) may be extended to more than two discrete random variables. Formally, discrete random variables Y_1, Y_2, \dots, Y_n are independent random variables if and only if the joint distribution of Y_1, Y_2, \dots, Y_n is equal to the product of their respective marginal distributions, that is:

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = f_{Y_1}(y_1) f_{Y_2}(y_2) \dots f_{Y_n}(y_n) \quad (6.20)$$

Similarly for the *cdf*s, discrete random variables Y_1, Y_2, \dots, Y_n are independent random variables if and only if the joint cumulative distribution of Y_1, Y_2, \dots, Y_n is equal to the product of their respective cumulative marginal distributions, that is:

$$F_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = F_{Y_1}(y_1) F_{Y_2}(y_2) \dots F_{Y_n}(y_n) \quad (6.21)$$

6.6 Distributions of continuous bivariate random variables

In parallel with distributions of discrete bivariate random variables, we have distributions of continuous bivariate random variables. The switch from discrete to continuous random variables is made by replacing summation by integration. The continuous counterpart of the discrete probability function is the probability density function (*pdf*) or density. The continuous counterpart of the cumulative distribution function for discrete bivariate random variables is also called a cumulative distribution function.

joint density Let X and Y be two continuous random variables whose densities are $f_X(x)$ and $f_Y(y)$, respectively. A **joint density** of X and Y is a function $f(x, y)$ that satisfies the following conditions:

$$f(x, y) \geq 0 \quad (6.22)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) = 1 \quad (6.23)$$

A graph of $f(x, y)$ would be three dimensional, with a surface of height $f(x, y)$ above the point (x, y) in the x - y plane.

marginal density The **marginal density** of X is obtained by integrating the joint density over y :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (6.24)$$

The marginal density of Y is obtained by integrating the joint density over x :

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (6.25)$$

conditional density The **conditional density** of Y given X is defined as:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad (6.26)$$

Similarly, the conditional density of X given Y is defined as:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \quad (6.27)$$

independent random variables

In parallel with the discussion of independent discrete random variables, two continuous random variables X and Y are **independent random variables**, or simply independent, if their joint density equals the product of the two marginal densities. That is,

$$f(x_i, y_j) = f_X(x_i) f_Y(y_j) \quad (6.28)$$

The *cdfs* of independent continuous random variables also multiply. That is,

$$F(x_i, y_j) = F_X(x_i) F_Y(y_j) \quad (6.29)$$

Both Equations (6.28) and (6.29) can be generalized: Y_1, Y_2, \dots, Y_n are independent random variables if and only if:

$$f(y_1, y_2, \dots, y_n) = f_1(y_1) f_2(y_2) \dots f_n(y_n) \quad (6.30)$$

or if:

$$F(y_1, y_2, \dots, y_n) = F_1(y_1) F_2(y_2) \dots F_n(y_n) \quad (6.31)$$

Throughout the remainder of this book, provided there is no ambiguity, we will use the term random variable for either a discrete or continuous random variable.

6.7 Expected value and variance

expected value

In studying the properties of a random variable Y , it is often convenient to evaluate one or more descriptive characteristics of Y . One key characteristic is the **expected value** of Y , denoted by $E[Y]$. If Y is discrete, the expected value of Y is the weighted average of the values taken on by Y , with the weights equal to their probabilities. If Y is discrete, its expected value

is calculated as:

$$E[Y] = \sum_i y_i Pr\{Y = y_i\} = \sum_i y_i f(y_i) \quad (6.32)$$

where the summation is over all values of Y .

mean expectation The expected value is sometimes called the **mean** or **expectation** and is often denoted by the Greek letter μ .

Another key characteristic of a random variable Y is its variance, denoted by $V[Y]$. The variance of Y is defined as the expected value of the square of the difference between Y and its mean μ , that is:

$$V[Y] = E[(Y - \mu)^2] \quad (6.33)$$

If Y is discrete, the variance is calculated as:

$$V[Y] = E[(Y - \mu)^2] = \sum_i (y_i - \mu)^2 Pr\{Y = y_i\} = \sum_i (y_i - \mu)^2 f(y_i) \quad (6.34)$$

➤ The variance is often denoted by σ^2 , the square of the Greek letter σ .

The standard deviation of a discrete random variable Y is defined as the positive square root of its variance and is usually denoted by σ . Thus,

$$\sigma = \sqrt{V(Y)} \quad (6.35)$$

Example 6.7 illustrates the expected value in a familiar setting.

Example 6.7. A Lottery. A lottery offers three prizes: \$250, \$100, and \$50. If 1000 tickets are sold at \$1 each, what is the expected value of a ticket?

Let Y represent the winnings with $y_1 = 0$, $y_2 = 50$, $y_3 = 100$, and $y_4 = 250$. The probability of not winning any prize is 997/1000, of winning \$50 is 1/1000, of winning \$100 is 1/1000, and of winning \$250 is 1/1000. Using Equation (6.32), the expected winnings for one ticket are:

$$E[Y] = (0)(0.997) + (50)(0.001) + (100)(0.001) + (250)(0.001) = 0.40.$$

If all 1000 tickets are sold, the total paid out in prizes is \$400, or \$0.40 per ticket sold. Thus, the expected value of the winnings is equal to the average paid out per ticket sold.

The variance of Y in this example is calculated using Equation (6.34) as:

$$\sigma^2 = (0 - 0.40)^2 (0.997) + (50 - 0.40)^2 (0.001) + (100 - 0.40)^2 (0.001) + (250 - 0.40)^2 (0.001) = 74.84$$

The standard deviation is $\sigma = 8.65$.

Whereas the variance and standard deviation appear to be irrelevant for this example, they may be important in a statistical analysis of the honesty and integrity of such lotteries.

expected value If Y is a continuous random variable, its **expected value**, $E[Y]$, or **mean**, μ , is defined analogously to the definition of the expected value of a discrete random variable in Equation (6.32). It is the integral of the values taken on by Y multiplied by its density function, where the integral is evaluated over the entire domain of Y . Formally,

$$E[Y] = \int y f(y) dy \quad (6.36)$$

domain The **domain** of Y is a set of values such that $f(y) > 0$. It is also called the **support** of Y .

If Y is a continuous random variable, its variance, $V[Y]$, is defined analogously to the definition of the variance of a discrete random variable in Equation (6.34). It is given by Equation (6.37), where $\mu = E[Y]$ and the integral is evaluated over the entire domain of Y .

$$V Y = \int (y - \mu)^2 f(y) dy \quad (6.37)$$

For both discrete and continuous random variables, it can be shown that the variance can also be calculated as:

$$V[Y] = E[Y^2] - (E[Y])^2 \quad (6.38)$$

The standard deviation of a continuous random variable is defined as the square root of its variance, the same as for a discrete random variable, and is also usually denoted by σ .

6.8 Linear combinations

linear combination A function $g = g(Y_1, Y_2, \dots, Y_n)$ of n random variables is called a **linear combination** if it can be written as:

$$g = g(Y_1, Y_2, \dots, Y_n) = c_0 + c_1 Y_1 + c_2 Y_2 + \dots + c_n Y_n \quad (6.39)$$

where c_0, c_1, \dots, c_n are constants.

If $g = g(Y_1, Y_2, \dots, Y_n)$ is a function of n random variables, g itself is a random variable. Theory shows that the mean $E[g]$ of a linear combination g can be written as follows:

$$E(g) = c_0 + c_1 E(Y_1) + c_2 E(Y_2) + \dots + c_n E(Y_n) \quad (6.40)$$

If the random variables Y_1, Y_2, \dots, Y_n are independent, then the variance $V[g]$ of a linear combination g can be written as follows:

$$V(g) = c_1^2 V(Y_1) + c_2^2 V(Y_2) + \dots + c_n^2 V(Y_n) \quad (6.41)$$

If the random variables are not independent, Equation (6.40) still holds but Equation (6.41) is incomplete. For example, if the random variables represent temperature readings on consecutive days, they will be correlated. To account for the lack of independence, Equation (6.41) must include one or more additional terms. The additional terms may be positive or negative, thus increasing or decreasing the value of $V[g]$. See Bowen and Bennett (1988), p. 89, for details.

random sample observations A key statistical concept is that of a **random sample** of n points or **observations**. A random sample is usually modeled by a set of independent random variables, each having the same distribution. In the remainder of this chapter, we describe some properties of the sample mean. The **sample mean** is the arithmetic mean of a random sample of n points such that:

- Repeated observations are drawn from a single random variable Y with mean μ and variance σ^2 .
- The sample observations are independent.

From Equation (6.39), the sample mean is a linear combination with $c_0 = 0$ and all other constants $c_1, c_2, \dots, c_n = 1/n$. Denoting the sample mean by \bar{Y} (read as Y bar), we have:

$$\bar{Y} = 0 + \frac{1}{n} Y_1 + \frac{1}{n} Y_2 + \dots + \frac{1}{n} Y_n \quad (6.42)$$

The sample mean \bar{Y} defined by this equation is a random variable. In practice, we calculate its value by replacing each Y_i in Equation (6.42) with its observed value y_i to get \bar{y} . We then use \bar{y} to make statistical inferences about the

population represented by Y . These tests are usually based on the expected value and variance of \bar{Y} .

From Equations (6.40) and (6.41), the expected value and variance of the sample mean \bar{Y} are:

$$E(\bar{Y}) = \frac{1}{n}\mu + \frac{1}{n}\mu + \dots + \frac{1}{n}\mu = \frac{1}{n} \sum_{i=1}^n \mu = \mu \quad (6.43)$$

$$\begin{aligned} V \bar{Y} &= \left(\frac{1}{n}\right)^2 \sigma^2 + \left(\frac{1}{n}\right)^2 \sigma^2 + \dots + \left(\frac{1}{n}\right)^2 \sigma^2 \\ &= n \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n} \end{aligned} \quad (6.44)$$

The standard deviation of \bar{Y} is:

$$\sigma_{\bar{Y}} = \sqrt{V(\bar{Y})} = \sqrt{\sigma^2/n} = \sigma/\sqrt{n} \quad (6.45)$$

standard error of the mean The standard deviation of the sample mean is also called the **standard error of the mean** or simply the **standard error**.

standard error

6.9 Linear contrasts

linear contrast contrast

A special case of a linear combination is when the sum of the coefficients $c_1 + \dots + c_n$ in Equation (6.39) is equal to 0. This linear combination is called a **linear contrast** or simply a **contrast**. A common example of a contrast is the difference between two independent sample means. From Equation (6.39), the contrast is:

$$g = \bar{Y}_1 - \bar{Y}_2 \quad (6.46)$$

where \bar{Y}_1 and \bar{Y}_2 are sample means based on sample size n_1 and n_2 . Here, $c_0 = 0$, $c_1 = 1$, and $c_2 = -1$.

From Equation (6.40), we have:

$$E(g) = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2 \quad (6.47)$$

From Equations (7.41) and (7.44), we have:

$$V(g) = V(\bar{Y}_1) + V(\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad (6.48)$$

and the standard deviation of g is:

$$\sigma_g = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (6.49)$$

standard error of mean differences The last expression is also called the **standard error of mean differences**.

An important special case of Equation (6.44) is when $n_1 = n_2 = 1$. We have

$$g = Y_1 - Y_2 \quad (6.50)$$

where Y_1 and Y_2 are two independent random variables. From Equation (6.46), the variance of g is

$$V(g) = \sigma_1^2 + \sigma_2^2 \quad (6.51)$$

In other words, the variance of the sum of two independent random variables is the sum of their variances. This result also follows from Equation (6.39) by setting $n = 2$, $c_0 = 0$, $c_1 = 1$, and $c_2 = -1$ in Equation (6.37).

6.10 Mean square error

mean square error (MSE) Measurements are often biased (see Section 5.3). The combination of known bias (from manufacturer's specs, for example) and sampling error is often expressed by the **mean square error**, also known as **MSE**, given in Equation (6.52). A discussion of mean square error can be found in Cochran (1977), p. 15.

$$\begin{aligned} MSE(\text{measurement}) = & \\ & (\text{variance of measurement}) \\ & + (\text{bias in measurement})^2 \end{aligned} \quad (6.52)$$

MSE is a measure of uncertainty that reflects both the bias and the random component of the measurement system.

7

Continuous distributions

7.1 What to look for in Chapter 7

In this chapter, we study selected continuous distributions and their associated densities and cumulative distribution functions (*cdfs*). We start with a simple distribution:

- *uniform distribution*, §7.2
- *standard uniform distribution*, §7.3

Next, we focus on the most commonly used continuous distribution:

- *normal distribution*, §7.4
- *standard normal distribution*, §7.5

Procedures for normal distribution probability calculations are shown using:

- *table lookup*, §7.6
- *spreadsheet routines*, §7.7

We continue with several distributions that are related to the normal distribution, which are needed later in the book:

- *sample mean distribution and the Central Limit Theorem, §7.8*
- *lognormal distribution, §7.9*
- *chi-square distribution, §7.10*
- *Student's t-distribution, §7.11*
- *F-distribution, §7.12*
- *exponential distribution, §7.13*
- *gamma distribution, §7.14*
- *beta distribution, §7.15*

For each distribution presented, the mathematical forms and graphs of its density and *cdf* are given along with its mean and variance. Also, where available, spreadsheet functions for the densities and *cdf*'s are shown along with examples of their use.

7.2 Uniform distribution

uniform distribution A *uniform distribution*, also called a *rectangular distribution*, models situations where a random variable Y takes on values only in a specified interval (a, b) such that Y falls in any two subintervals of (a, b) of equal length with the same probability.

rectangular distribution

The uniform distribution is denoted by $U(a, b)$. Its density is:

$$f(y) = \begin{cases} \frac{1}{b-a}, & a \leq y \leq b \\ 0, & \text{otherwise} \end{cases} \quad (7.1)$$

Figure 7.1 shows a graph of the density of $U(a, b)$. The dashed vertical lines at $y = a$ and $y = b$ indicate that $f(y)$ is discontinuous at these points.

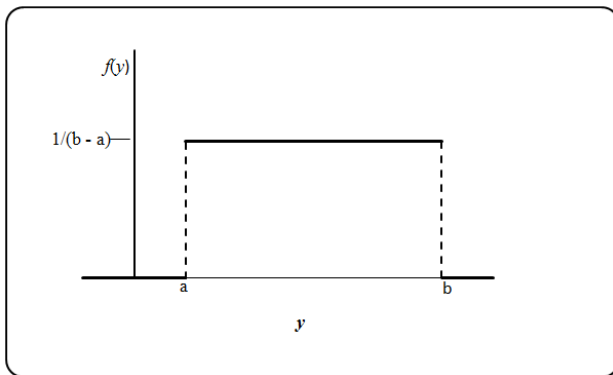


Figure 7.1. Density function of the uniform distribution $U(a, b)$

The *cdf* of $U(a, b)$ is calculated using Equation (6.7) as:

$$\begin{aligned} F(y) &= 0, & y < a \\ &= \frac{y-a}{b-a}, & a \leq y \leq b \\ &= 1, & b < y \end{aligned} \quad (7.2)$$

Suppose that I is a subinterval of (a, b) with length $w < b - a$. Then, from Equations (6.8) and (7.2), the probability that a random variable Y from $U(a, b)$ falls in I is:

$$Pr\{Y \text{ in } (a, b)\} = \frac{w}{b-a} \quad (7.3)$$

The *cdf* for $U(a, b)$ is graphed in Figure 7.2.

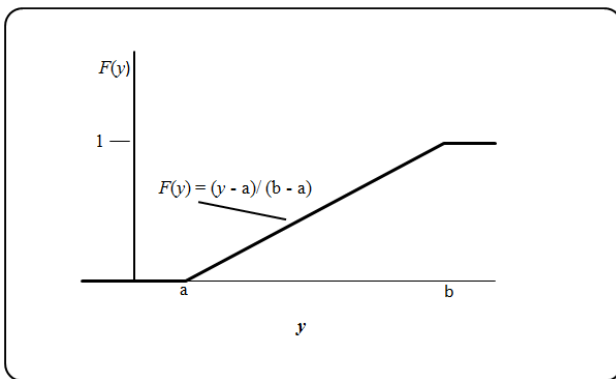


Figure 7.2. cdf of a uniform distribution

To obtain the mean, variance, and standard deviation of $Y = U(a, b)$, substitute the density from Equation (7.2) into Equations (6.36) and (6.37). As in Section 6.7, we use the Greek letter notation to denote these parameters.

$$\mu = E(Y) = \frac{(b+a)}{2} \quad (7.4)$$

$$\sigma^2 = V(Y) = \frac{(b-a)^2}{12} \quad (7.5)$$

$$\sigma = \frac{b-a}{\sqrt{12}} \quad (7.6)$$

Example 7.1. Fuel pump gauge. A pump at the gas station is read to the nearest 0.01 of a gallon. The error Y associated with the reading is a rounding error which is uniformly distributed between $a = -0.005$ and $b = 0.005$. The mean, variance, and standard deviation of Y , respectively, are calculated as:

$$\mu = \frac{0.005 + (-0.005)}{2} = 0$$

$$\sigma^2 = \frac{(0.005 - (-0.005))^2}{12} = 0.00000833$$

$$\sigma = 0.00289$$

7.3 Standard uniform distribution

standard uniform distribution Of special interest is $U(0, 1)$, the uniform distribution on the interval $(0, 1)$, called the **standard uniform distribution** or simply the standard uniform. Its density and *cdf* are:

$$f(y) = 1, 0 \leq y \leq 1 \quad (7.7)$$

$$F(y) = y \quad (7.8)$$

The mean, variance, and standard deviation of $Y = U(0, 1)$ are:

$$\mu = 0.5 \quad (7.9)$$

$$\sigma^2 = \frac{(1-0)^2}{12} = \frac{1}{12} = 0.0833 \quad (7.10)$$

$$\sigma = \sqrt{0.0833} = 0.289 \quad (7.11)$$

An important use of the standard uniform distribution is to generate random numbers from any distribution for which the *cdf* is explicitly known or well approximated. Details are given in Chapter 27, "Simulations."

Additional information about the uniform distribution may be found in the following texts:

Krishnamoorthy (2006), p. 115

Hastings and Peacock (1975), p. 116

Patel et al. (1976), p. 33

7.4 Normal distribution

normal distribution The **normal distribution** plays a prominent role in statistical analyses. The importance of the normal distribution cannot be overstated because, apart from its mathematical tractability, it provides a model for many natural phenomena and supports a wide variety of decisionmaking processes. Moreover, many phenomena and measurement systems that do not follow the normal distribution can be transformed into an approximate normal distribution by relatively simple mathematical manipulations.

The modern mathematical formulation of the normal distribution is derived from a function developed by Carl Friedrich Gauss (1777–1855) in his 1809 book *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* (*Theory of the Motion of the Heavenly Bodies Revolving Around the Sun in Conic Sections*).

Gaussian distribution Consequently, some authors refer to the normal as the **Gaussian distribution**, after its developer.

normal density The **normal density** is given by:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}}, \quad -\infty < y < \infty \quad (7.12)$$

The two parameters shown in Equation (7.12) are μ and σ^2 , where μ is the mean and σ^2 is the variance of the normal distribution. A plot of the normal density is shown in Figure 7.3, where the familiar “bell-shaped” curve centered about the mean μ is recognized.

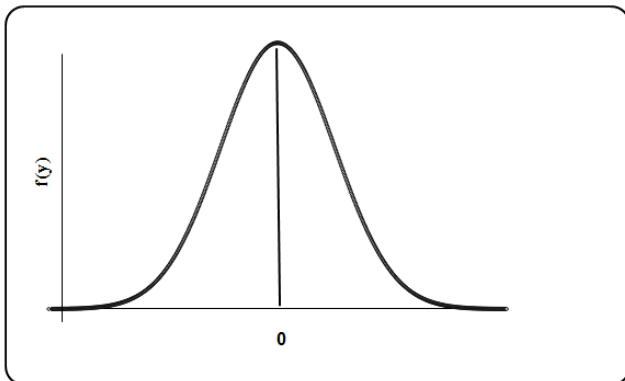


Figure 7.3. Normal density function

In examining Figure 7.3, the following characteristics of the normal density are apparent. These are characteristics that some other, nonnormal densities may also have:

- The symmetry of the curve about μ means that observations from a normal distribution are as likely to occur above μ as below μ .
- The decrease in the curve as we move away from the center means that large deviations from μ are less likely to occur than are small deviations.

It is important to note that not every bell-shaped density is normal. Indeed, throughout this text, we will study other distributions that are similar to the normal distribution. In performing statistical procedures, it is important to distinguish between a normal distribution and a similar but different distribution. Accordingly, there are statistical procedures that test the validity of the assumption of normality. See Chapter 11, “Goodness-of-Fit Tests,” for details.

For descriptive and statistical purposes, datasets are often modeled by a normal distribution. The following example presents a dataset where normality is plausible.

Example 7.2. U.S. natural uranium concentration (1993 map, downloaded from U.S. Geological Survey site). Figure 7.4 presents a pictorial dataset that appears to be normal. In this figure, every dot, or pixel, has an associated value, typically between 0 and 6 ppm, of natural uranium concentration. Because not every point in the continental United States is represented in Figure 7.4, the values associated with Figure 7.4 are only a sample from the population of natural uranium concentrations in the continental United States. We make no claim that the dataset is a random sample that could be used for statistical investigations. For this example, the dataset is used only for descriptive purposes.

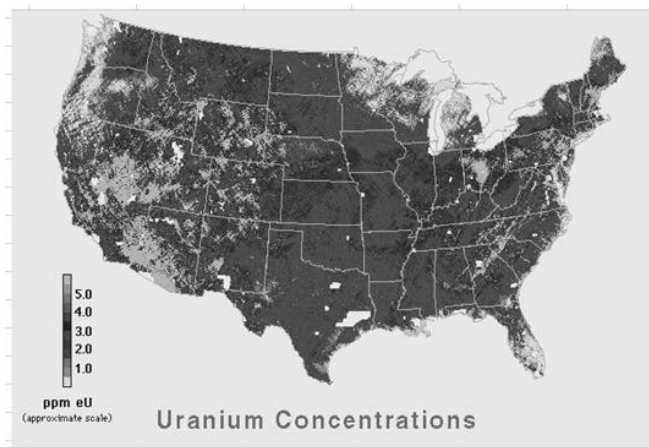


Figure 7.4. Natural uranium concentration
 Source: U.S. Geological Survey: Digital Data Series DDS-9

If the data in this example are arranged in ascending order, we could sketch a histogram similar to that shown in Figure 7.5. This histogram (when properly scaled) is analogous to a density function. Because Figure 7.5 is approximately symmetric and bell-shaped, it is plausible to conjecture that the dataset of uranium concentrations in that figure is normally distributed. See Chapter 11 for a discussion of statistical procedures that could be used to test this conjecture.

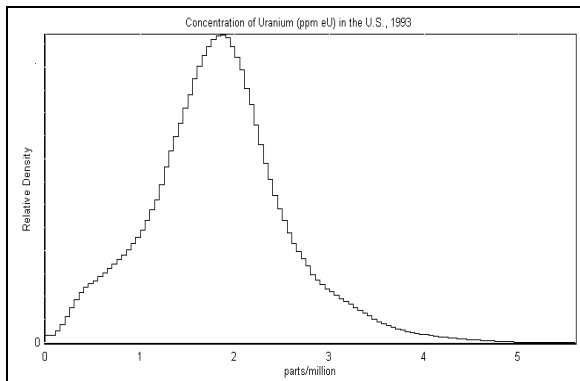


Figure 7.5. Histogram of natural uranium concentration

The *cdf* for the normal distribution is given by $F(y)$, where:

$$F(y) = Pr\{Y \leq y\} = \int_{-\infty}^y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(w-\mu)^2}{\sigma^2}} dw, \quad -\infty < y < \infty \quad (7.13)$$

The conventional notation used for the normal distribution is $N(\mu, \sigma^2)$. A shorthand notation such as $Y \sim N(\mu, \sigma^2)$ is read as “ Y is distributed normally with a mean μ and variance σ^2 .”

- Note that the second parameter, σ^2 , is the variance, not the standard deviation. For example, the standard deviation of $N(10, 25)$ is 5, not 25.
- The normal distribution has the useful property that any linear function of a normal distribution is also normal, although the mean and/or variance might change. Specifically, if $Y \sim N(\mu, \sigma^2)$ and $U = aY + b$, where a and b are any constants, then:

$$U \sim N(a\mu + b, a^2\sigma^2) \quad (7.14)$$

7.5 Standard normal distribution

standard normal distribution Let $Y \sim N(\mu, \sigma^2)$. Of particular importance is the **standard normal distribution**, commonly denoted by Z , defined as:

$$Z = \frac{Y - \mu}{\sigma} \quad (7.15)$$

To find the mean and variance of Z , set $a = 1/\sigma$ and $b = -\mu/\sigma$ in Equation (7.14) to get:

$$Z \sim N(0, 1) \quad (7.16)$$

In other words, the mean of a standard normal is always 0 and its variance is always 1. That is why it is called a standard normal.

The density of Z is obtained from Equation (7.12) by setting $\mu = 0$ and $\sigma = 1$. The resultant density is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty \quad (7.17)$$

The *cdf* for the standard normal is commonly denoted by $\Phi(z)$ where Φ is the upper case Greek letter phi. Setting $\mu = 0$ and $\sigma = 1$ in Equation (7.13), we have:

$$\Phi(z) = Pr\{Z \leq z\} = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw, \quad -\infty < z < \infty \quad (7.18)$$

- Some texts use the notation $N(z; 0, 1)$ or, simply, $N(z)$, rather than $\Phi(z)$, to denote the cumulative standard normal distribution.

Figure 7.6 shows the graph of $\Phi(z)$.

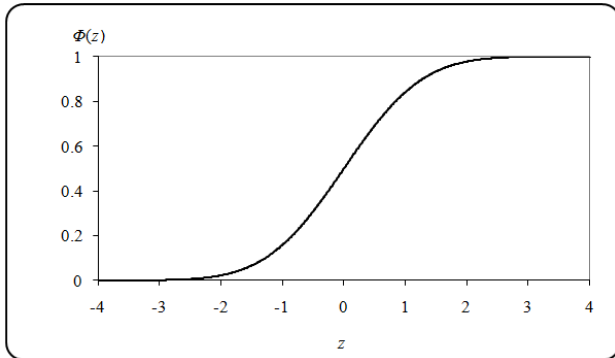


Figure 7.6. Standard normal CDF

In probabilistic and statistical analyses of normally distributed variables, we often need to calculate probabilities associated with a specific quantile or to calculate the quantile associated with a specific probability. In either case, we need to evaluate integrals in the form of Equation (7.13). Unfortunately, that integral cannot be evaluated in closed form (i.e., in terms of standard mathematical functions). However, there are two approaches to evaluating a normal *cdf*. One of these approaches uses a table lookup, while another approach resorts to a spreadsheet subroutine call. Section 7.6 discusses table lookup steps, and Section 7.7 addresses spreadsheet function calls.

7.6 Table lookup for the normal distribution

In principle, a table of the *cdf* for the normal distribution $Y \sim N(\mu, \sigma^2)$ should allow us to determine (with interpolation as necessary) the probability that $Y \leq y$ for just about any specific quantile. Unfortunately, it would be impossible to create such a table for every combination of μ and σ that would be needed for applications. However, because of the special relation between an arbitrary normal Y and the standard normal Z as expressed in Equation (7.15), the *cdf* of any normal Y can be evaluated in terms of the *cdf* $\Phi(z)$ of Z .

For example, suppose we wish to evaluate $F(b)$, where $F(y)$ is the *cdf* of $Y \sim N(\mu, \sigma^2)$. From Equation (7.13), we have $F(b) = Pr\{Y \leq b\}$. Using the definition of the standard normal from Equation (7.15), we see that:

$$F(b) = Pr\{Y \leq b\} = Pr\left\{\frac{Y - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right\} = Pr\left\{Z \leq \frac{b - \mu}{\sigma}\right\} \quad (7.19)$$

From the definition of $\Phi(z)$ in Equation (7.18), we have:

$$F(b) = \Phi((b - \mu)/\sigma) \quad (7.20)$$

standard normal table Thus, Equation (7.19) provides a formula for the *cdf* of Y in terms of the *cdf* of Z . Table T-1, “The cumulative standard normal distribution,” of the appendix gives the values of $\Phi(z)$. This table is also called a **standard normal table** or a **z-table**.

We notice that Table T-1 covers z values up to $z = 3.49$. This should not present a problem because the probability of z exceeding 3.49 is less than 0.0002, a negligible amount for most applications. If $\Phi(z)$ is needed for $z > 3.49$, extended tables or approximations are available and could be used for this purpose. Also, a spreadsheet routine call could be used, as shown in Section 7.7.

In obtaining probability values from a standard normal table, bear in mind the following properties:

- The probability that Z is less than or equal to a value a is:

$$Pr\{Z \leq a\} = \Phi(a) \quad (7.21)$$

- The probability that Z is larger than a value b is:

$$Pr\{Z > b\} = 1 - Pr\{Z \leq b\} = 1 - \Phi(b) \quad (7.22)$$

The probability that Z is between two values a and b , where $a < b$, is:

$$\begin{aligned} Pr\{a < Z < b\} &= Pr\{a < Z \leq b\} = Pr\{a \leq Z < b\} \\ &= Pr\{a \leq Z \leq b\} = \Phi(b) - \Phi(a) \end{aligned} \quad (7.23)$$

Equation (7.23) follows from the fact that $Pr\{Y = y\} = 0$ for any continuous random variable Y (see Equation (6.9)).

By the symmetry of the normal distribution about its mean, it also has the following property:

- The probability that Z is less than a value a , where $a < 0$ is equal to 1 minus the probability that Z is greater than a . That is:

$$Pr\{Z < a\} = 1 - Pr\{Z > -a\} = 1 - \Phi(-a) \quad (7.24)$$

Thus, although a standard normal extends from $-\infty$ to $+\infty$, only positive values of z are needed in seeking probabilities in the standard normal table.

To make these probability calculations more concrete, turn now to Table T-1. To find the probability that corresponds to a given positive z value, look up the first two leading digits of z in the left column (table entries showing from 0.0 to 3.4), and then look up the next digit as a column heading. The desired probability is found at that row and column intersection. As an example, the probability that corresponds to $z = 1.23$ is found at the intersection of the row listed as 1.2 and the column listed as 0.03. The table value reads 0.8907, which is the probability that $Z \leq 1.23$. If we require the probability that corresponds to $z = 1.234$, then interpolation between 0.8907 (for which $z = 1.23$) and 0.8925 (for which $z = 1.24$) would be required. A linear interpolation would be $0.8907 + (1.234 - 1.230)(0.8925 - 0.8907)/(1.24 - 1.23) = 0.8914$.

Before use of Table T-1, making a rough sketch of the normal density is often very valuable. That sketch would indicate which area of the distribution should be added and which subtracted to obtain the desired probability. The following example shows such sketches.

Example 7.3. UO₂ enrichment. An extensive history of powder-batch measurements supports the assumption that the stoichiometric factor for percent of uranium dioxide (UO₂) powder manufactured at a given facility is normally distributed with a mean $\mu = 87.5$ and a standard deviation $\sigma = 0.4$. Suppose that a sample of a single measurement is to be taken.

- What is the probability that this measurement will be less than or equal to 88.0?

$$Pr\{Y \leq 88.0\} = Pr\left\{\frac{Y-87.5}{0.4} \leq \frac{88.0-87.5}{0.4}\right\} = Pr\{Z \leq 1.25\} = \Phi(1.25) = 0.8944$$

The probability area of interest is sketched in the hashed part of Figure 7.7 for the original normal Y and in Figure 7.8 for the standard normal Z .

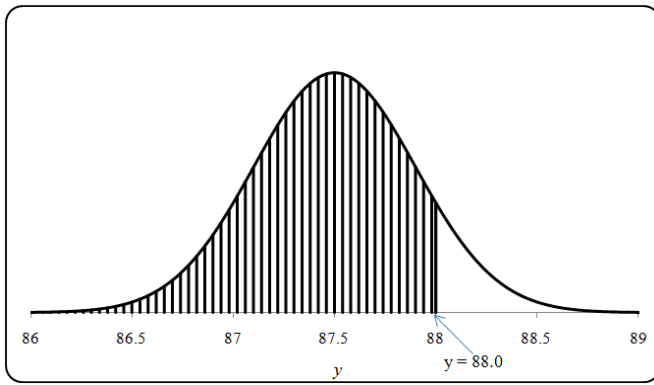


Figure 7.7. Sketch of $Pr\{Y \leq 88.0\}$

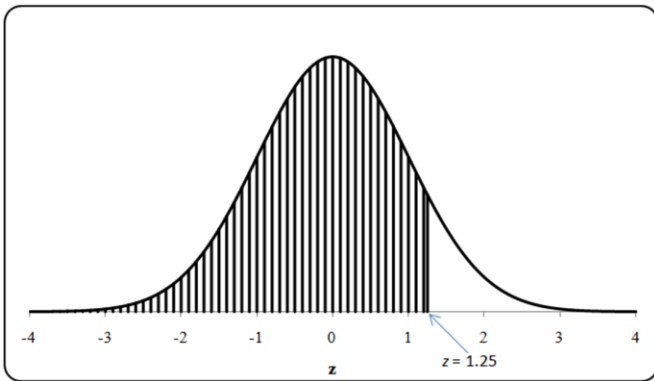


Figure 7.8. Sketch of $Pr\{Z \leq 1.25\}$

- What is the probability that this measurement will be larger than 88.0?

$$\begin{aligned} Pr\{Y > 88.0\} &= Pr\left\{\frac{Y - 87.5}{0.4} > \frac{88.0 - 87.5}{0.4}\right\} = Pr\{Z > 1.25\} \\ &= 1 - Pr\{Z < 1.25\} = 1 - 0.8944 = 0.1056 \end{aligned}$$

The probability area of interest is sketched in the hashed part of Figure 7.9 in terms of Z only.

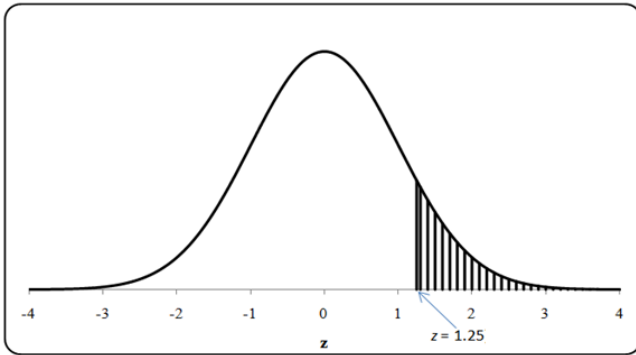


Figure 7.9. Sketch of $Pr\{Z > 1.25\}$

- What is the probability that this measurement will be between 87.0 and 88.1?

$$\begin{aligned} Pr\{87.0 < Y \leq 88.1\} &= Pr\left\{\frac{87.0-87.5}{0.4} < Z \leq \frac{88.1-87.5}{0.4}\right\} = Pr\{-1.25 < Z \leq 1.50\} \\ &= Pr\{Z \leq 1.50\} - Pr\{Z \leq -1.25\} = 0.9332 - (1 - 0.8944) \\ &= 0.9332 - 0.1056 = 0.8276 \end{aligned}$$

The probability area of interest is sketched in the shaded part of Figure 7.10.

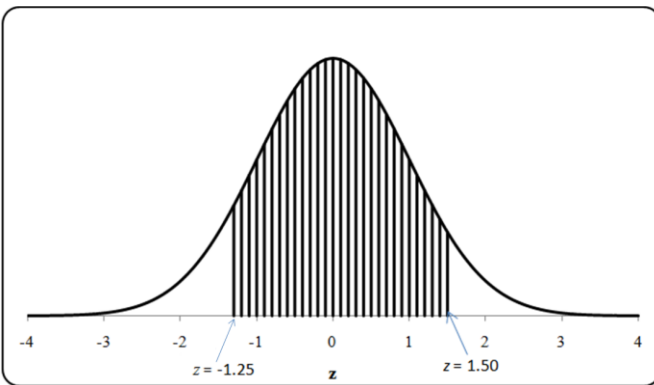


Figure 7.10. Sketch of $Pr\{-1.25 < Z \leq 1.50\}$

7.7 Spreadsheet functions for normal probabilities

Much of the tedium of evaluating an integral such as Equation (7.18), or even the effort of table lookup and interpolation, could be alleviated by using a spreadsheet call. Excel provides three functions for this purpose:

=NORMDIST, =NORMINV, and =NORMSINV. We explain the use of these functions below.

=NORMDIST($y, \mu, \sigma, 1$) returns the probability that a normal random variable Y with a mean μ and standard deviation σ will be as large as y . The constant “1” is an indicator that tells the function that a cumulative normal (rather than a density) is desired. Graphically, =NORMDIST($y, \mu, \sigma, 1$) returns the area under the curve for which $-\infty < Y < y$, as sketched in Figure 7.7.

As an example of the use of the =NORMDIST function, we show the following values obtained by this function in support of the calculations in Example 7.3.

- =NORMDIST(88.0, 87.5, 0.4, 1) = 0.8944
- =NORMDIST(87.0, 87.5, 0.4, 1) = 0.1056
- =NORMDIST(88.1, 87.5, 0.4, 1) = 0.9332

These values are the same as those obtained from Table T-1. Discrepancies between Table T-1 values and Excel’s value may occur if Table T-1 is interpolated. Generally, such discrepancies are inconsequential.

- If the last entry in the =NORMDIST is given as 0 rather than 1, the density of the distribution (height of the curve) is returned. Using the 0 entry is useful in plotting the normal density, as was done for Figure 7.5.

=NORMINV(p, μ, σ) is Excel’s function that performs the inverse operation of =NORMDIST. =NORMINV(p, μ, σ) returns a quantile q so that p (or $100p\%$) of the area under the $N(\mu, \sigma^2)$ distribution curve is less than q . For example, to find a value y such that 95% of the distribution of $N(75, 0.16)$ is below y , a call to =NORMINV(0.95, 75.0, 0.4) yields $y = 75.658$.

=NORMSINV (note the letter “S” in =NORMSINV which stands for “standard”) performs a function similar to that of =NORMINV for the standard normal distribution. In this case, the parameters μ and σ are not needed as they are assumed to be 0 and 1, respectively. Accordingly, the call to this function is =NORMSINV(p) and the only function input is the probability p such as =NORMSINV(p) or ($100p\%$). An example of a call to this function is =NORMSINV(0.95), which returns the value of $q_{0.95} = 1.645$.

The spreadsheet normal inverse functions =NORMINV and =NORMSINV are used extensively in statistical tests of hypotheses. For this purpose, Table 7.1 lists commonly used quantiles along with the values returned by Excel’s =NORMINV and =NORMSINV functions. These values can also be found for selected probabilities at the bottom of Table T-1 in the appendix to this book.

Table 7.1. Selected values for the standard normal distribution

Quantile, q	Excel's standard normal inverse function	Excel's general normal inverse function	z value
0.005	=NORMSINV(0.005)	=NORMINV(0.005, 0, 1)	-2.576
0.01	=NORMSINV(0.01)	=NORMINV(0.01, 0, 1)	-2.326
0.025	=NORMSINV(0.025)	=NORMINV(0.025, 0, 1)	-1.960
0.05	=NORMSINV(0.05)	=NORMINV(0.05, 0, 1)	-1.645
0.01	=NORMSINV(0.01)	=NORMINV(0.01, 0, 1)	-1.282
0.02	=NORMSINV(0.02)	=NORMINV(0.02, 0, 1)	-0.842
0.05	=NORMSINV(0.5)	=NORMINV(0.5, 0, 1)	0.000
0.80	=NORMSINV(0.8)	=NORMINV(0.8, 0, 1)	0.842
0.90	=NORMSINV(0.9)	=NORMINV(0.9, 0, 1)	1.282
0.95	=NORMSINV(0.95)	=NORMINV(0.95, 0, 1)	1.645
0.975	=NORMSINV(0.975)	=NORMINV(0.975, 0, 1)	1.960
0.99	=NORMSINV(0.99)	=NORMINV(0.99, 0, 1)	2.326
0.999	=NORMSINV(0.995)	=NORMINV(0.995, 0, 1)	2.576

Note that the 0.025 and the 0.975 quantiles of the standard normal distribution are -1.96 and 1.96, respectively. This means that 95% of the $N(\mu, \sigma^2)$ lies between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$. This is consistent with the empirical rule of Chapter 2 that claims that about 95% of a mound-shaped distribution lies within 2 standard deviations of the mean. We can also verify, using Table T-1 or Excel =NORMSINV, that nearly 2/3 of the distribution lies between $\mu - \sigma$ and $\mu + \sigma$ and that nearly the entire distribution lies between $\mu - 3\sigma$ and $\mu + 3\sigma$.

7.8 The Central Limit Theorem

Section 6.8 introduced the key statistical concepts of a random sample and a sample mean. If a random sample of size n is drawn from a random variable Y with mean μ and standard deviation σ , the sample mean \bar{Y} has mean μ and standard deviation σ/\sqrt{n} . If μ is unknown, as is usually the case, the sample mean can then be used to estimate μ . However, to make statistical statements about the uncertainty of this estimate, it is also necessary to know something about the distribution of \bar{Y} . The problem is that the distribution of \bar{Y} is determined by the distribution of Y , which is usually unknown. A commonly used resolution to this dilemma stems from a remarkable property of the normal

distribution. In many cases of interest, the distribution of \bar{Y} turns out to be approximately normal. Thus, for a large class of possible distributions of Y , statistical inferences about μ can be made from a knowledge of \bar{Y} and the known properties of the normal distribution.

Central Limit Theorem (CLT) The theoretical basis for this approach is called the **Central Limit Theorem** or **CLT**. When its conditions are met, this theorem provides powerful ways of answering a variety of statistical questions.

There are several versions of the *CLT*. Here we focus on one version, which reads:

Suppose that a random sample of n observations is drawn from a distribution of a random variable Y with mean μ and standard deviation σ , satisfying certain regulatory conditions. Then the distribution of the sample mean \bar{Y} is asymptotically normal with mean μ and standard deviation σ/\sqrt{n} . Mathematically, this means that if $Y \sim ?(\mu, \sigma^2)$ then :

$$\bar{Y} \rightarrow N(\mu, \sigma^2/n) \text{ as } n \rightarrow \infty \quad (7.25)$$

where the question mark indicates “any distribution” and the symbol \rightarrow means “converges to” or “approaches.”

To summarize, here are the conditions and implications of the *CLT*:

- A random sample of size n from a distribution of a random variable Y is observed. The parameters of Y and its distribution are not necessarily known.
- Assume that the distribution of Y has a mean μ and a standard deviation σ .
- As the sample size, n , increases, the distribution of the sample mean \bar{Y} approaches that of the normal distribution.
- The mean of \bar{Y} is μ , the same mean as that of the original distribution Y .
- The standard deviation of \bar{Y} is σ/\sqrt{n} .

The benefits of the *CLT* are three-fold:

- The sample mean, on the average, is the same as the distribution mean.
- The distribution—or at least the approximate distribution—of the sample mean is now known.
- The standard deviation (standard error) $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$ of \bar{Y} is smaller (by a factor of \sqrt{n}) than the standard deviation of Y .

Equation (7.25) can also be written in terms of the standard normal variable Z .

$$\text{If } Y \sim \mathcal{N}(\mu, \sigma^2) \text{ then } Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1) \text{ as } n \rightarrow \infty \tag{7.26}$$

A proof of this version of the *CLT* stated above may be found in Mood et al. (1974), p. 195.

The key conclusion of the *CLT* is that the distribution of the sample mean is asymptotically normal. The key assumption in its proof is that the sample mean is essentially a sum of identical independent random variables. Although the *CLT* as stated above applies only to the sample mean, other versions of the *CLT* apply to other sums of independent random variables, which may not even be identical. Because real world measurements can often be modeled as sums of independent components, this explains why so many measurements are, at least approximately, normally distributed.

Another version of the *CLT* applies the sum of independent and identical random variables, as follows:

Suppose that a random sample of n points is drawn from a random variable with a mean μ and a standard deviation σ . Then the distribution of the sample sum $\sum Y = n\bar{Y}$ is asymptotically normal with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$.

Table 7.2 lists the basic parameters of the sample mean \bar{Y} and sum $\sum Y$.

Table 7.2. Parameters of the sample mean and sample sum

Variable	Sample size	Mean	Variance	Standard deviation
Y	1	μ	σ^2	σ
\bar{Y}	n	μ	σ^2/n	σ/\sqrt{n}
$\sum Y = n\bar{Y}$	n	$n\mu$	$n\sigma^2$	$\sigma\sqrt{n}$

The *CLT* states that the sample mean is asymptotically normal (i.e., its distribution approaches a normal distribution as the sample size n approaches infinity). The practical question remains as to how large n has to be in order for the approximate normality to be useful. Here are some rough guidelines:

According to Hines and Montgomery (1980), p. 183:

If observations come from a near-normal distribution (near bell-shaped;

symmetric about the mean; tapering away from the mean) a sample as small as 4 may be adequate.

If observations come from a flat (uniform) continuous distribution, a sample of size 12 is adequate. In fact, early computers generated random normal observations by repeatedly taking averages of 12 uniform observations from the $U(0,1)$ distribution.

Also, according to Hald (1952a), p. 201:

In many cases it has been found that for $n > 30$, the distribution of the sample mean is fairly normal and even for smaller values of n the normal distribution is often sufficiently accurate.

The sample sizes cited above may be used to decide whether we are close enough to normality. However, considerations other than sample size may sometimes take precedence. Sample size determination for some cases will be discussed later in this book.

7.9 Lognormal distribution

lognormal distribution The use of the **lognormal distribution** has become increasingly widespread. It is commonly used as a distribution for failure frequency and in maintainability analysis. It has also been widely used as a prior distribution for unknown positive parameters in Bayesian analyses. See Chapter 20 for details.

A random variable Y is defined to be lognormal if the natural logarithm of Y is normal. Equivalently, Y is lognormal if:

$$Y = e^W, \text{ where } W \sim N(\mu, \sigma^2) \quad (7.27)$$

underlying normal We write $Y \sim LN(\mu, \sigma^2)$. The normal W is sometimes called the **underlying normal** of the lognormal Y .

The density of a lognormal random variable Y is:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(y)-\mu]^2}, \quad 0 < y < \infty, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0 \quad (7.28)$$

The mean, variance, median, and mode of a lognormal distribution are, respectively:

$$E[Y] = e^{\mu + \frac{\sigma^2}{2}} \quad (7.29)$$

$$V[Y] = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \quad (7.30)$$

$$\text{Median}[Y] = e^{\mu} \quad (7.31)$$

$$\text{Mode}[Y] = e^{\mu - \sigma^2} \quad (7.32)$$

Sometimes e^{μ} , the median of Y , is used as a parameter in the description of the lognormal distribution, so be vigilant.

The mode indicates the position of the peak of the density.

To calculate probabilities for $Y = \ln(\mu, \sigma^2)$, the table for the standard normal can be used. Specifically, for any number a ,

$$\begin{aligned} Pr\{Y \leq a\} &= Pr\{\ln(Y) \leq \ln(a)\} \\ &= Pr\left\{Z \leq \frac{\ln(a) - \mu}{\sigma}\right\} \end{aligned} \quad (7.33)$$

where Z is the standard normal.

Because $Y = e^W$ is an increasing function of W , the quantiles of Y correspond to the quantiles of W . Thus, if q_p is the p^{th} quantile of W , then e^{q_p} is the p^{th} quantile of Y . In particular, because μ is the median of W , it follows that e^{μ} is the median of Y .

error factor A commonly used parameter of a lognormal Y is the **error factor (EF)**, defined as the ratio of the 0.95 quantile (95th percentile) of Y to its median. Because of the logarithmic symmetry of the lognormal, the error factor is also equal to the ratio of the median to the 0.05 quantile. As shown in Table T-1 of the appendix, the 0.95 quantile of the underlying normal W is $\mu + 1.645\sigma$. From Equation (7.27), the 0.95 quantile of Y is $e^{\mu + 1.645\sigma}$. Because e^{μ} is the median of Y , we have:

$$EF = e^{1.645\sigma} \quad (7.34)$$

Note that the error factor does not depend on μ .

Figure 7.11 shows three lognormal densities. The value $\mu = -7$ corresponds to a median of e^{-7} , or about 0.0009. The value $\mu = -6.5$ corresponds to a median of about 1.5×10^{-3} . The value $\sigma = 1.4$ corresponds to an error factor $EF = e^{1.6457\sigma} = e^{1.645(1.4)} = 10$ and $\sigma = 0.67$ corresponds to an error factor $EF = e^{1.645(0.67)} = 3$.

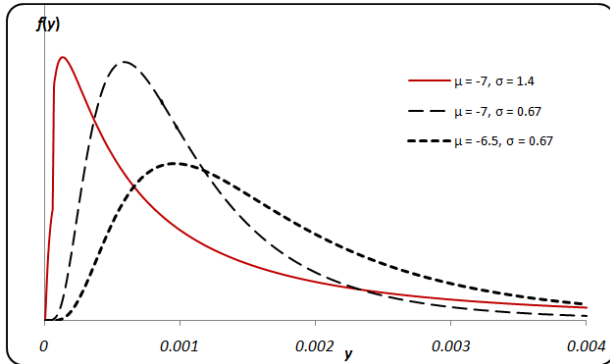


Figure 7.11. Three lognormal densities

The two lognormals with $\sigma = 0.67$ and different values of μ but the same $EF = 3$ have essentially the same shape, but with different scales. The larger μ corresponds to spreading the distribution away from zero. The lognormal with $\sigma = 1.4$ and $EF = 10$ is much more skewed than the lognormal with $EF = 3$.

The lognormal distribution often arises from the product of several independent random variables. If Y is the product of n independent positive random variables that are identically (or nearly so) distributed, then:

$$\ln(Y) = \ln((Y_1)(Y_2) \dots (Y_n)) = \ln \prod_{i=1}^n Y_i = \sum_{i=1}^n \ln(Y_i) \quad (7.35)$$

is approximately normal because of the *CLT* property of sums of independent and identical random variables. Consequently, Y is approximately lognormal.

Use Excel's spreadsheet function =NORMDIST($\ln(y)$, μ , σ , 1) to calculate $F(y)$. Alternatively, use Excel's spreadsheet function =LOGNORMDIST(y , μ , σ) to calculate $F(y)$. Note that the function =NORMDIST requires $\ln(y)$, whereas =LOGNORMDIST requires only y . The input terms for the =LOGNORMDIST function are:

y = the value at which to calculate the *cdf*

μ = the mean of $\ln(Y)$

σ = the standard deviation of $\ln(Y)$

As examples, we have the following functions and their returns:

LOGNORMDIST(0.001, -7, 4) = 0.526

NORMDIST(ln(0.001), -7, 4, 1) = NORMDIST(-6.91, -7, 4, 1) = 0.526

where $\ln(0.001) = -6.91$

Additional information about the lognormal distribution may be found in the following texts:

Krishnamoorthy (2006), p. 247

Hastings and Peacock (1975), p. 84

Patel et al. (1976), p. 28

7.10 Chi-square distribution

chi-square distribution The *chi-square distribution* is a widely used distribution because, under reasonable assumptions, many statistics follow that distribution. Specific uses of the chi-square distribution include goodness-of-fit tests (Chapter 11), contingency table analyses (Chapter 12), and tests of variance (Chapter 14). This section defines and characterizes the distribution; other chapters demonstrate and explain the applications.

degrees of freedom The chi-square distribution is denoted by $\chi^2(\nu)$, where the symbol χ is the lower case Greek letter chi and the parameter ν is the lower case Greek letter nu. In statistical applications, ν is a positive integer and is also known as *degrees of freedom*.

The density of a random variable Y that has the chi-square distribution with ν degrees of freedom is given by:

$$f(y) = \frac{e^{-y/2} y^{(\nu-2)/2}}{2^{(\nu/2)} \Gamma(\nu/2)}, \quad y > 0, \quad \nu = 1, 2, 3, \dots \quad (7.36)$$

The symbol Γ in Equation (7.36) is the Greek capital letter gamma. It is the standard notation for the gamma function that is defined for $w > 0$ as:

$$\Gamma w = \int_0^{\infty} x^{w-1} e^{-x} dx \quad (7.37)$$

➤ For a positive integer k , $\Gamma(k) = (k-1)! = (1)(2) \dots (k-1)$. Note that $\Gamma(1) = 0! = 1$.

- The chi-square distribution with ν degrees of freedom can also be defined as the sum of squares of ν independent and identically distributed standard normal random variables. Mathematically,

$$Y = \sum_{i=1}^{\nu} Z_i^2 \quad (7.38)$$

where Z_i indicates independent standard normal variables.

The density of the chi-square distribution is plotted in Figure 7.11 for selected values of ν .

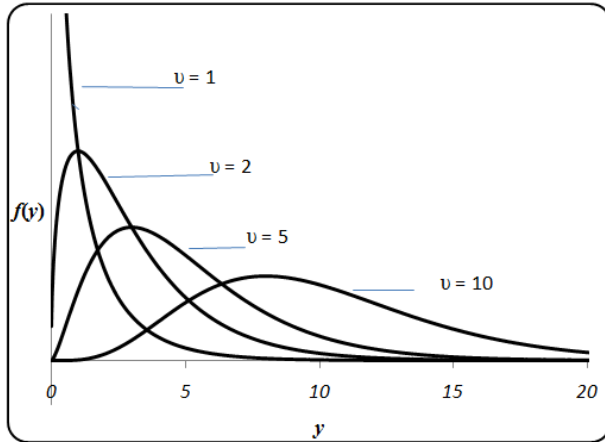


Figure 7.12. Chi-square density for 1, 2, 5, and 10 degrees of freedom

The mean and variance of Y are simple functions of ν .

$$E[Y] = \nu \quad (7.39)$$

$$V[Y] = 2\nu \quad (7.40)$$

Selected quantiles of the chi-square distribution for selected degrees of freedom are given in Table T-2, “Quantiles, $\chi_q^2(\nu)$, for the chi-square distribution with ν degrees of freedom,” of the appendix. The q^{th} quantile of Y , denoted by y_q , is a value such that $100q\%$ of the $\chi^2(\nu)$ distribution is below y_q and $(1-q)100\%$ of the distribution is above. y_q . To find y_q , look at Table T-2 under the row for ν and the column for q . As an example, Table T-2 shows the 0.95 quantile $\chi_{0.95}^2(5)$ of the chi-squared distribution for 5 degrees of freedom as 11.1.

Figure 7.13 shows this graphically. The shaded area under the curve for $0 < y < 11.1$ represents 95% of the chi-square distribution with $\nu = 5$ degrees of freedom. Accordingly, the probability that $\chi^2(5)$ will exceed 11.1 is only 0.05, or 5%, as reflected in the unshaded area to the right of $y = 11.1$.

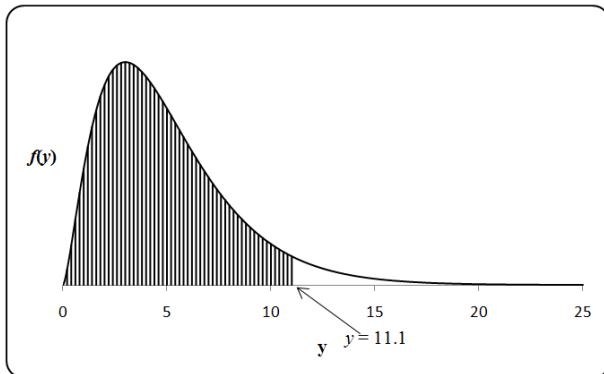


Figure 7.13. 95% cumulative probability for the $\chi^2(5)$ distribution

As an alternative to table lookup, you can use a spreadsheet function that, for a selected probability, returns the associated quantile. Excel's spreadsheet function `=CHIINV(1-p, ν)` returns the value y_0 that is associated with a probability q that a random variable $\chi^2(\nu)$ will be less than y_0 . For example, the call `=CHIINV(0.05, 5)`, for the 95th quantile of the $\chi^2(\nu)$ distribution, returns the value of $y_0 = 11.1$.

A related spreadsheet function is Excel's `=CHIDIST(y, ν)` which returns the probability $\Pr\{Y > y_0\}$ for a given value of $y = y_0$. For example, `=CHIDIST(11.1, 5)` returns the value of 0.05.

Note that Excel does not give the density of the chi-square distribution. For that matter, the density is not very useful for data analysis. However, if we wish to calculate that chi-square density (perhaps for plotting the distribution), we can calculate it directly from Equation (7.36).

- To calculate the density of $\chi^2(\nu)$ from Equation (7.36), it is necessary to evaluate the gamma function in Equation (7.37). Excel provides the function `=GAMMALN`, the natural logarithm of the gamma function, but not the gamma function itself. Thus, to calculate $\Gamma(w)$, use a combination of Excel's functions such as `=EXP(GAMMALN(w))`, where `=EXP()` is Excel's natural antilogarithm function.

Additional information about the chi-square distribution may be found in the following texts:

Johnson and Kotz (1970), p. 166

Hastings and Peacock (1975), p. 124

Krishnamoorthy (2006), p. 155

7.11 Student's t -distribution

t -distribution Student's **t -distribution** is a widely used distribution because, under reasonable assumptions, many statistics follow that distribution. Specific uses of the t -distribution include estimation techniques (Chapter 9) and testing hypotheses about population parameters (Chapters 13, 15, and others). This section defines and characterizes the distribution; other chapters demonstrate and explain the applications.

Student's t -distribution The mathematical foundation of **Student's t -distribution** or simply, the **t -distribution** was laid in 1908 in a publication by W.S. Gosset, who served as a brewmeister with the Guinness brewery in Dublin. The brewery forbade research publications by its staff (because an earlier paper contained trade secrets), and consequently, Gosset published his work under the pseudonym of "Student."

The mathematical definition of the t -distribution is as follows. Suppose Z has a standard normal distribution, X has a chi-square distribution with ν degrees of freedom, and Z and X are independent. Then,

$$Y = \frac{Z}{X/\nu} \quad (7.41)$$

has a t -distribution with ν degrees of freedom, where ν is a positive integer.

The t -distribution is denoted by $t(\nu)$. Its density is given by:

$$f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < y < \infty \quad (7.42)$$

where Γ is the gamma function from Equation (7.37). The density of $t(\nu)$ is symmetric about 0, and, as ν increases, approaches the density of the standard normal distribution.

The mean of Y is:

$$E[Y] = 0, \quad \nu = 2, 3, \dots \tag{7.43}$$

➤ The mean of Y is not defined when $\nu = 1$.

The variance of Y is:

$$V[Y] = \frac{\nu}{\nu - 2}, \quad \nu = 3, 4, \dots \tag{7.44}$$

➤ The variance of Y is not defined when $\nu = 1$ or $\nu = 2$.

Figure 7.14 shows the density of Y for $\nu = 1, 4,$ and 100 . As ν increases, the density converges quickly to the standard normal density. Indeed, if a graph of the standard normal density were superimposed on Figure 7.14, it would be indistinguishable from the t -distribution density for $\nu = 100$. In fact, it would be essentially indistinguishable for $\nu = 30$.

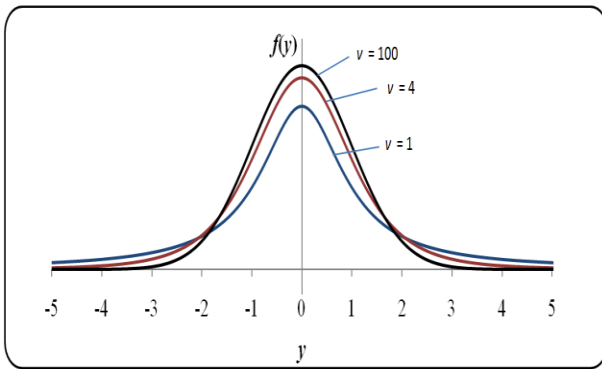


Figure 7.14. Student’s t density for $\nu = 1, 4,$ and 100

Selected quantiles q of the t -distribution for selected degrees of freedom ν are given in Table T-3, “Quantiles, $t_q(\nu)$, for Student’s t -distribution with ν degrees of freedom.” of the appendix. The q^{th} quantile of $t(\nu)$, denoted by $t_q(\nu)$, is a value such that $100q\%$ of the $t(\nu)$ distribution is below $t_q(\nu)$ and $(1 - q)100\%$ of the distribution is above $t_q(\nu)$. To find a specific quantile q , look at Table T-3 under the row for ν and the column for q . As an example, Table T-3 shows the 0.975 quantile $t_{0.975}(30)$ of the t -distribution for 30 degrees of freedom as 2.04. This is shown graphically in Figure 7.15, where the shaded area to the left of 2.04 represents 97.5% of the distribution, and the unshaded area greater than 2.04 represents 2.5% probability of the Student’s t -distribution with $\nu = 30$ degrees of freedom.

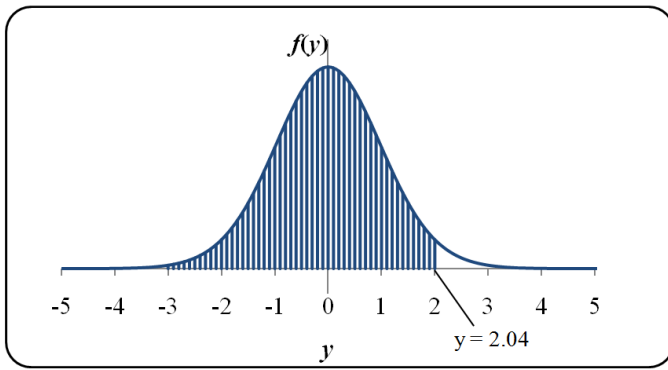


Figure 7.15. 97.5% cumulative probability for the t -distribution with $\nu = 30$

Just as for the normal distribution, the t -distribution is symmetric about its mean of 0. Thus, Table T-3 shows only positive quantiles (i.e., only for values of $q \geq 0.50$). If $q < 0.50$, the required quantile is the negative of the $(1 - q)^{\text{th}}$ quantile (i.e., $-t_{1-q}(\nu)$). Thus, the 0.025 quantile of $t(30)$ is $-t_{0.975}(30) = -2.04$.

As an alternative to table lookup, use a spreadsheet function that, for a selected probability q , yields the associated quantile. Excel's `=TINV` function returns the complement quantile, but assumes that a two-sided interval is requested. Hence, to get $t_q(\nu)$, the call to Excel is `=TINV(2(1 - q), ν)`. Thus, to find the 0.975 quantile of $t(30)$, call `=TINV(0.05, 30)` and obtain the value of 2.04 (rounded).

A less practical spreadsheet function is Excel's `=TDIST(y_0 , ν , k)`. When $k = 2$, this function returns the probability of Y exceeding a point y_0 . When $k = 1$, this function returns half that probability. Examples of the use of this function are `=TDIST(2.04, 30, 2)` which returns a value of 0.05, and `=TDIST(2.04, 30, 1)` which returns a value of 0.025.

Additional information about the t -distribution may be found in the following texts:

Hastings and Peacock (1975), p. 120

Krishnamoorthy (2006), p. 172

7.12 *F*-distribution

***F*-distribution** The ***F*-distribution** is a widely used distribution because, under reasonable assumptions, many statistics follow that distribution. Specific uses of the *F*-distribution include tests of equality of means of several distributions (Chapters 16 and 17), tests of equality of variances (Chapter 14), and equality of regression lines (Chapter 18). This section defines and characterizes the distribution; other chapters demonstrate and explain the applications.

The *F*-distribution was originally defined by Sir Roland A. Fisher in 1920 as the distribution of the ratio of two independent chi-square variables. George E. Snedecor later modified this distribution as the ratio of two independent chi-square variables, each divided by its degrees of freedom. Formally, Snedecor’s *F* (so named to honor Fisher’s contribution) is defined as:

$$F = \frac{\chi_1^2/v_1}{\chi_2^2/v_2} \tag{7.45}$$

where χ_1^2 and χ_2^2 are independent chi-square variables and v_1 and v_2 are the corresponding degrees of freedom.

The *F*-distribution is denoted by $F(v_1, v_2)$. The parameters v_1 and v_2 are usually referred to as degrees of freedom of the numerator and degrees of freedom of the denominator, respectively.

A distinguishing property of the *F*-distribution is that the reciprocal of an *F*-distribution is also an *F*-distribution. From Equation (7.45), we have:

$$1/F_{v_1, v_2} = \left[\frac{\chi_1^2/v_1}{\chi_2^2/v_2} \right]^{-1} = \frac{\chi_2^2/v_2}{\chi_1^2/v_1} = F_{v_2, v_1} \tag{7.46}$$

For v_1 and v_2 positive integers, the density function of the *F*-distribution is:

$$f(y) = \frac{\left(\frac{v_1}{v_2}\right)^{v_1/2}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} y^{(v_1-2)/2} \left(1 + \frac{v_1}{v_2} y\right)^{-\left(\frac{v_1+v_2}{2}\right)}, \quad y > 0 \tag{7.47}$$

where the denominator is the beta function, defined for positive a and b by:

$$B(a, b) = \int_0^1 w^{a-1} (1-w)^{b-1} dw \quad (7.48)$$

The mean of the F -distribution, defined for $\nu_2 > 2$, is:

$$E[Y] = \frac{\nu_2}{\nu_2 - 2}, \quad \nu_2 = 3, 4, \dots \quad (7.49)$$

The variance of the F -distribution, defined for $\nu_2 > 4$, is:

$$V[Y] = \frac{2\nu_2^2}{\nu_1} \frac{\nu_1 + \nu_2 - 2}{\nu_2 - 2} \frac{1}{\nu_2 - 4}, \quad \nu_2 = 5, 6, \dots \quad (7.50)$$

Figure 7.16 shows the density of the F -distribution for several combinations of ν_1 and ν_2 . Figure 7.16 is reproduced from *NIST Handbook* (2006), Section 1.3.6.6.5.

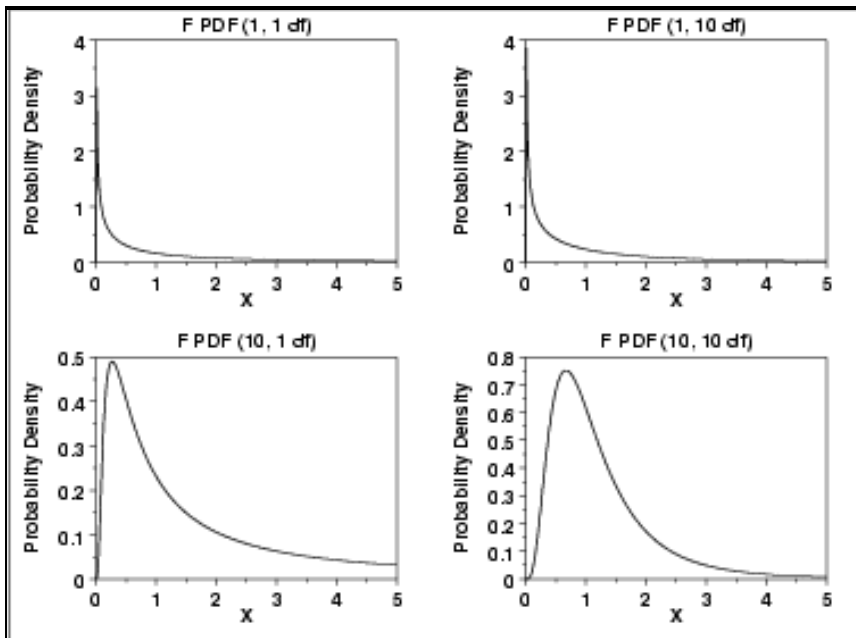


Figure 7.16. Density function of the F -distribution for selected combinations of degrees of freedom

Table T-4, “Quantiles, $f_q(\nu_1, \nu_2)$, for the F -distribution,” of the appendix gives selected quantiles of $F(\nu_1, \nu_2)$ for various combinations of ν_1 and ν_2 . To find a quantile, follow these steps:

- Locate the row that corresponds to the degrees of freedom of the denominator, v_2 .
- Locate the column that corresponds to the degrees of freedom of the numerator, v_1 .
- Locate the sub-row that corresponds to the required quantile.

As an example, the 0.95 quantile of the F -distribution with $v_1 = 5$ and $v_2 = 20$ is $f_{0.95}(5,20) = 2.71$.

Because of space limitations, Table T-4 provides only commonly used quantiles. For the same reason, only selected degrees of freedom for the numerator and denominator are listed. If necessary, use interpolation, or better yet, an appropriate spreadsheet function, as discussed later.

Note that Table T-4 lists only quantiles $q \geq 0.90$. To obtain quantiles $q \leq 0.10$, we exploit the reciprocal property of the F -distribution as expressed by Equation (7.46).

Let Y be any distribution that takes on only positive values, and let y_q be the q^{th} quantile of Y . Then,

$$q = Pr\{Y < y_q\} = Pr\{Y > 1/y_q\} = 1 - Pr\{1/Y < 1/y_q\} \quad (7.51)$$

From Equation (7.50), $Pr\{1/Y < 1/y_q\} = 1 - q$. In other words, $1/y_q$ is the $(1-q)^{\text{th}}$ quantile of $1/Y$. Now set $Y = F(v_2, v_1)$ and let $f_q(v_2, v_1)$ be its q^{th} quantile. Then, $1/f_q(v_2, v_1)$ is the $(1-q)^{\text{th}}$ quantile of $1/F(v_2, v_1) = F(v_1, v_2)$, using Equation (7.46). We conclude that the $(1-q)^{\text{th}}$ quantile of $F(v_1, v_2)$ is $1/f_q(v_2, v_1)$. We write:

$$f_{(1-q)}(v_1, v_2) = \frac{1}{f_q(v_2, v_1)} \quad (7.52)$$

Note the reversal of the degrees of freedom from Equation (7.51).

As an example, to find the 0.05 quantile of the F -distribution with $v_1 = 5$ and $v_2 = 20$, we first look up the table value for the 0.95 quantile, $f_{0.95}(20,5) = 4.56$. Then the reciprocal of that value, $1/4.56 = 0.219$, is the 0.05 quantile of $F(5,20)$.

Excel provides two spreadsheet functions that are related to the F -distribution: =FDIST and =FINV.

Excel's =FDIST function may be used to calculate a complementary cumulative probability—the probability that a random observation from the F -distribution will exceed a specified value, y_0 . The call to this function is =FDIST(y_0, v_1, v_2), where v_1 is the degrees of freedom in the numerator and v_2 is the degrees of freedom in the denominator. Examples of the use of the =FDIST function are:

$$=FDIST(9.45, 24, 2) = 0.100$$

$$=FDIST(1.80, 4, 30) = 0.155$$

$$=FDIST(1.80, 30, 4) = 0.303$$

The first example states that the probability of a random observation from $F(24, 2)$ exceeding 9.45 is 0.100. The value of the *cdf* at $y_0 = 9.45$ is $1 - 0.100 = 0.900$.

The last two examples of the =FDIST function show that if the degrees of freedom are interchanged, different results are returned.

Excel's spreadsheet function =FINV determines the quantile whose probability of exceedance is a specified value p_0 . This function is used to obtain values such as in Table T-4 for any combination of degrees of freedom and p_0 and with more significant figures.

The call to this function is =FINV (p_0, v_1, v_2), where:

p_0 is the probability for which the value of F is calculated

v_1 is the degrees of freedom in the numerator

v_2 is the degrees of freedom in the denominator

Examples of the use of the =FDIST function are:

$$=FINV(0.100, 24, 2) = 9.45$$

$$=FINV(0.05, 1, 100) = 3.94$$

Note that the =FINV's returned value is given from the complementary distribution of F . Hence, to obtain the q^{th} quantile, the first entry in this function must be the complementary probability (i.e., $1 - p_0$).

Additional information about the F -distribution may be found in the following texts:

Hastings and Peacock (1975), p. 124

Krishnamoorthy (2006), p. 202

7.13 Exponential distribution

exponential distribution The *exponential distribution* is widely used for modeling time to failure of a component or system. Denote the time to failure by the random variable $Y = Y(\lambda)$, where $\lambda > 0$ is a parameter.

The density of Y is given by:

$$f(y) = \lambda e^{-\lambda y}, y > 0 \quad (7.53)$$

The *cdf* of the exponential distribution is:

$$F(y) = 1 - e^{-\lambda y} \quad (7.54)$$

The mean and variance of T are:

$$E[Y] = 1/\lambda \quad (7.55)$$

$$V[Y] = 1/\lambda^2 \quad (7.56)$$

failure frequency Here, λ is the *failure frequency* and is equal to $1/E(Y)$. Therefore, λ has units of failures per unit time.

Figure 7.17 shows two exponential densities for two values of λ . The intercept (height of the curve when $y = 0$) equals λ . Thus, the figure shows that the distribution is more concentrated near zero if λ is large. This agrees with the interpretation of λ as a failure frequency and t as time to failure.

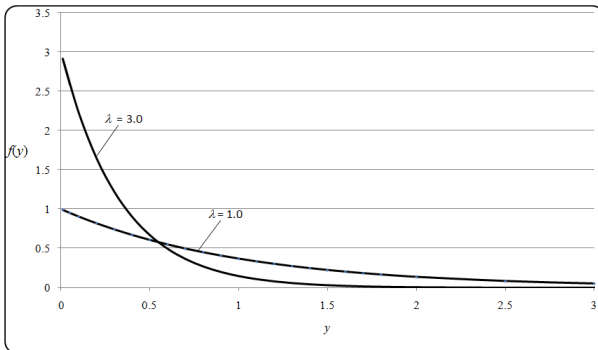


Figure 7.17. Two exponential densities

failure rate The exponential parameter λ is referred to as the **failure rate** if the component or system is repaired and restarted immediately after each failure. It is called the **hazard rate** if the component or system can fail only once and cannot be repaired. The hazard rate is defined as modeling duration times with an arbitrary density $f(y)$:

$$h(y) = \frac{f(y)}{1 - F(y)} \quad (7.57)$$

From Equations (7.52) and (7.53), the exponential hazard rate is constant and is equal to λ .

A different and sometimes useful parameterization of the exponential uses $\mu = 1/\lambda = E[Y]$. If Y is the time to failure, μ is called the **mean time to failure**. If Y is the time to repair, or to fire suppression, or to some other event, μ is called the **mean time to repair**, or another appropriate name. With parameter μ , the density of Y is given by:

$$f(y) = \frac{1}{\mu} e^{-y/\mu}, \quad y > 0 \quad (7.58)$$

and the *cdf* is:

$$F(y) = 1 - e^{-y/\mu}, \quad y > 0 \quad (7.59)$$

The units of μ are the same units of time as the data. The mean and variance are:

$$E[Y] = \mu \quad (7.60)$$

$$V[Y] = \mu^2 \quad (7.61)$$

Additional information about the exponential distribution may be found in the following texts:

Hastings and Peacock (1975), p. 56

Krishnamoorthy (2006), p. 179

Patel et al. (1976), p. 212

7.14 Gamma distribution

The gamma distribution is a generalization of both the exponential and chi-square distributions. Also, the sum of independent exponential random variables has a gamma distribution. The gamma distribution has a number of useful applications, including confidence intervals for exponential data, waiting time to the n^{th} event in a Poisson process, and a failure time model (Martz and Waller, 1991, Section 4.2). It is also often used as a prior distribution in Bayesian estimation (Chapter 20) of the failure rate parameter from Poisson or exponential data.

The density of a gamma random variable Y is:

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta}, \quad y > 0, \quad \alpha > 0, \quad \beta > 0 \quad (7.62)$$

where $\Gamma(\alpha)$ is defined in Equation (7.37) but is reproduced here for convenience:

$$\Gamma w = \int_0^{\infty} x^{w-1} e^{-x} dx$$

A gamma distribution with parameters α and β is denoted by $\text{Gamma}(\alpha, \beta)$. Its mean and variance are:

$$E[Y] = \alpha / \beta \quad (7.63)$$

$$\text{Var}[Y] = \alpha / \beta^2 \quad (7.64)$$

The parameters α and β are referred to as the shape and scale parameters, respectively. By varying α we change the appearance of the density. If α is near zero, the distribution is highly skewed. When $\alpha = 1$, the gamma distribution reduces to an exponential distribution.

Figure 7.18 shows gamma densities with four shape parameters, α . When $\alpha < 1$, the density becomes infinite at 0. When $\alpha = 1$, the density is identical to an exponential density. When α is large, the distribution is approximately normal.

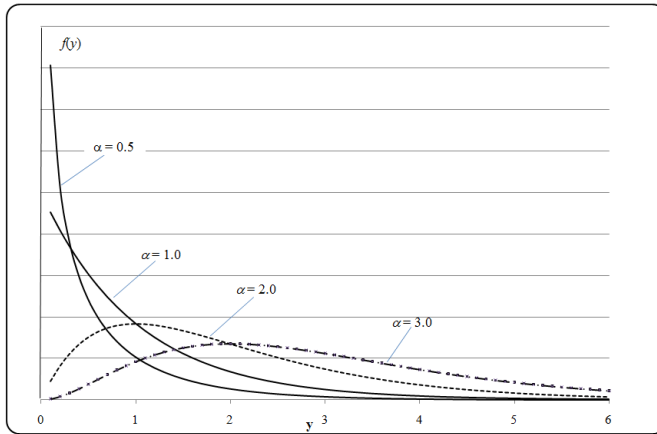


Figure 7.18 Gamma density distributions for $\beta = 1$ with four shape parameters.

Excel's spreadsheet function for gamma(α , β) is `GAMMADIST(y, α , β , 1)`, where y is the value at which we want to evaluate the distribution. The last term in parenthesis of this function is 1, instructing the spreadsheet to return the *cdf* evaluated at y . In contrast, the function `=GAMMADIST(y, α , β , 0)`, returns the density at y , which is not very useful.

Example: To obtain the probability that a gamma random variable will not be larger than $y = 105$ when $\alpha = 20$ and $\beta = 100$ is

$$=\text{GAMMADIST}(105, 20, 100, 1) = 0.393.$$

Additional information about the gamma distribution may be found in the following texts:

Hastings and Peacock (1975), p. 68,

Patel *et al* (1976), p. 30.

7.15 Beta Distribution

Many continuous quantitative phenomena take on values that are bounded by known numbers a and b . Examples are percentages, proportions, and ratios. The beta distribution is a versatile family of distributions that is useful for modeling phenomena that can range from 0 to 1 and, through a transformation, from a to b .

The beta distribution family includes the uniform distribution and density shapes that range from decreasing to uni-modal right-skewed to symmetric to U-shaped

to uni-modal left-skewed to increasing (Martz and Waller 1991, p. 103). It can serve as a reliability model for the probability that a system or component would operate for at least t units of time. The beta distribution is also widely used in Bayesian estimation and reliability analysis as a prior distribution for the binomial parameter p that represents a reliability or failure probability (see Chapter 20).

The density of a beta random variable Y is:

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 \leq y \leq 1, \quad \alpha > 0, \quad \beta > 0 \quad (7.65)$$

A beta distribution with parameters α and β is denoted by $Beta(\alpha, \beta)$. Its mean and variance are:

$$E[Y] = \frac{\alpha}{\alpha + \beta} \quad (7.66)$$

$$V[Y] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (7.67)$$

Various beta densities are shown in Figures 7.19 and 7.20. Figure 7.19 shows beta densities with $\alpha = \beta$, and therefore with mean 0.5. When $\alpha < 1$, the density becomes infinite when $y = 0$, and when $\beta < 1$, the density becomes infinite when $y = 1$. When $\alpha = \beta = 1$, the beta density reduces to the uniform. When $\alpha = \beta = 0.5$, the density is U shaped and is the Jeffreys noninformative prior for a binomial likelihood function (see Chapter 20). When α and β are large, the density is approximately normal.

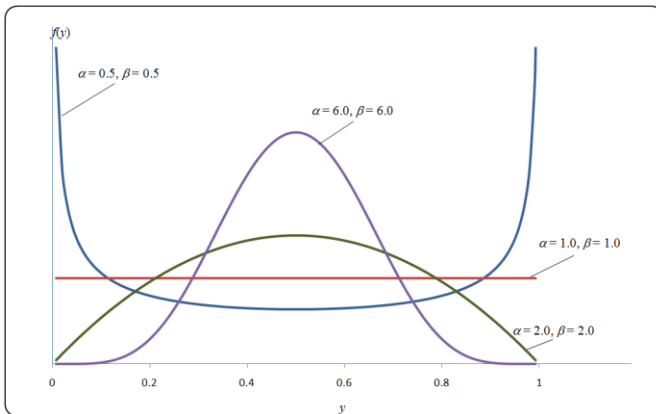


Figure 7.19. Four beta distributions with mean 0.5

Figure 7.20 shows four densities with mean 0.1. Again, when $\alpha < 1$, the density becomes infinite at 0, and when $\alpha > 1$, the density is zero at 0. When α and β are large, the density is approximately normal.

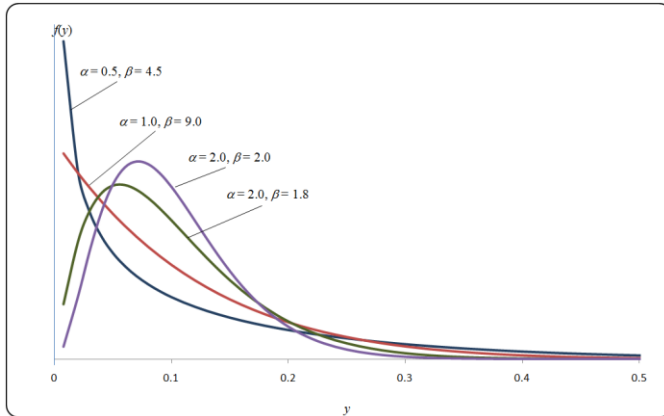


Figure 7.20. Four beta distributions with mean 0.1

Another parameterization of the beta distribution uses the parameters $x_0 = \alpha$ and $n_0 = \alpha + \beta$. This parameterization is used by Martz and Waller (1991) because it simplifies the formulas for Bayesian estimation. The density of a $Beta(x_0, n_0)$ distribution is

$$f(y) = \frac{\Gamma(n_0)}{\Gamma(x_0)\Gamma(n_0 - x_0)} y^{x_0 - 1} (1 - y)^{n_0 - x_0 - 1}, \quad 0 \leq y \leq 1, \quad 0 < x_0 < n_0 \quad (7.68)$$

The mean and variance of a $Beta(x_0, n_0)$ distribution are

$$E[Y] = x_0 / n_0 \quad (7.69)$$

$$V[Y] = \frac{x_0(n_0 - x_0)}{n_0^2(n_0 + 1)} \quad (7.70)$$

Excel's call function for the *cdf* of $Beta(\alpha, \beta)$ is =BETADIST(y, α, β, a, b), which returns the probability that $Y < y$ and where a and b are general bounds on Y . For the standard beta with boundaries 0 and 1, the call function is =BETADIST($y, \alpha, \beta, 0, 1$). Some examples for beta(2,18) are:

$$=BETADIST(0.1, 2, 18, 0, 1) = 0.580$$

$$=BETADIST(0, 2, 18, 0, 1) = 0.0$$

$$=BETADIST(1, 2, 18, 0, 1) = 1$$

The last two examples are trivial but may be used as a check on our function setup.

To illustrate the generalized distribution, suppose the random variable $Beta(2,18)$ was divided by 10 and its boundaries shifted from $(0, 1)$ to $(0.1, 0.2)$. Then $y = 0.1$ is transformed to $y = 0.11$ and the corresponding point on the *cdf* is $=BETADIST(0.11, 2, 18, 0.1, 0.2) = 0.580$, the same probability as for the standard beta.

Additional information about the beta distribution may be found in the following texts:

Hastings and Peacock (1975), p. 70,

Patel *et al* (1976), p. 216.

8

Discrete distributions

8.1 What to look for in Chapter 8

*discrete
distribution*

Chapter 8 lays the background for several discrete distributions. A *discrete distribution* is a distribution that applies to experiments where the set of values that can be observed is finite or countably infinite.

The distributions covered in this chapter are:

- *discrete uniform distribution*, §8.2
- *Bernoulli distribution*, §8.5
- *hypergeometric distribution*, §8.6
- *binomial distribution*, §8.7
- *geometric distribution*, §8.8
- *negative binomial distribution*, §8.9
- *Poisson distribution*, §8.10

probability function This chapter presents several discrete distributions. For each distribution, we give the **probability function** or **pf** and the associated mean, variance, and standard deviation, and point to typical situations where these distributions are applicable. We also give some spreadsheet functions that ease the calculations associated with these distributions.

probability mass function Recall that the probability function for discrete variables is sometimes called the **probability mass function** or **pmf** rather than **pf**.

Because of their importance, three of the listed distributions also have their own chapters with extensive discussion of their applications. These distributions are the hypergeometric distribution (Chapter 21), the binomial distribution (Chapter 22), and the Poisson distribution (Chapter 23).

As background for the Bernoulli and hypergeometric distributions, we also discuss:

- *sampling for attributes*, §8.3
- *sampling with and without replacement*, §8.4

8.2 Discrete uniform distribution

discrete uniform distribution The **discrete uniform distribution** is a probability distribution taking on a finite number of equally probable values. If a random variable Y may attain any of n possible values y_1, y_2, \dots, y_n , each with probability $1/n$, then Y is said to have a discrete uniform distribution. The **pf** for Y is given by:

$$f(y; n) = 1/n, y = y_1, y_2, \dots, y_n \quad (8.1)$$

The n possible values are often the integers $1, 2, \dots, n$. For example, if we need to randomly select a single individual from a class of size n , we can assign the members of the class consecutive numbers 1 through n and then draw (from a hat or otherwise) a random number between 1 and n to select one individual.

Example 8.1. Die rolling. When a fair die is rolled, each face is equally likely to turn up. Let Y equal the number of dots that turn up. Then Y has a discrete uniform distribution with **pf**:

$$f(y; 6) = 1/6, y = 1, 2, 3, 4, 5, 6 \quad (8.2)$$

If the n possible outcomes are in $[a, b]$, that is, starting at a number a and ending in b , we have $n = b - a + 1$ consecutive numbers, each with equal probability of being selected. The *pf* for this discrete uniform distribution is given by:

$$f(y; b - a + 1) = 1/(b - a + 1), y = a, a+1, \dots, b \quad (8.3)$$

Example 8.2. Travel vouchers. The travel office numbers travel vouchers consecutively, starting with 1 at the beginning of the year. Suppose we wish to randomly select a voucher from the month of March, where the vouchers are numbered from 392 through 516. From Equation (8.3), $a = 392$, $b = 516$, $n = (516 - 392 + 1) = 125$, and the probability that any specific voucher will be selected is $1/125 = 0.008$.

Let Y be a discrete uniform distribution assuming values $1, 2, \dots, n$. Its mean, variance, and standard deviation are:

$$\mu = \frac{(n+1)}{2} \quad (8.4)$$

$$\sigma^2 = (n+1)(n-1)/12 \quad (8.5)$$

$$\sigma = \sqrt{(n+1)(n-1)/12} \quad (8.6)$$

If Y is a discrete uniform distribution assuming $n = b - a + 1$ values in $[a, b]$, its mean is different than in Equation (8.4), but its variance and standard deviation are the same as in Equations (8.5) and (8.6):

$$\mu = (a + b)/2 \quad (8.7)$$

$$\sigma^2 = (b - a + 2)(b - a)/12 = (n + 1)(n - 1)/12 \quad (8.8)$$

$$\sigma = \sqrt{(b - a + 2)(b - a)/12} = \sqrt{(n + 1)(n - 1)/12} \quad (8.9)$$

The graph of the *pf* is the familiar bar chart of Section 3.5. It is plotted in Figure 8.1 for Example 8.1 (die rolling).

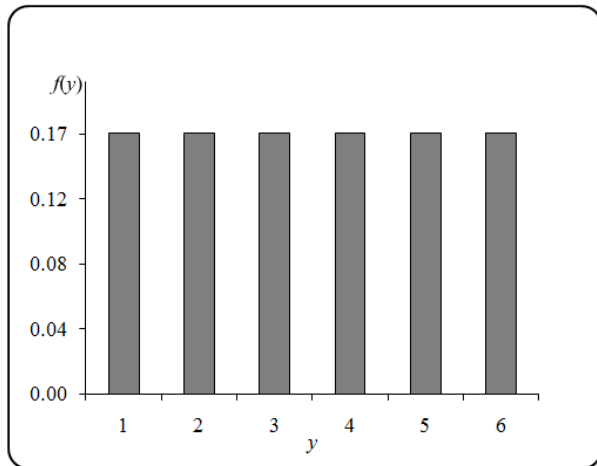


Figure 8.1. Probability function for die rolling

**discrete
random
number**

A **discrete random number** is defined as an observation from a discrete uniform distribution. In practice, there are several ways to select a random number. To select a random number from among n values, we write the n values on n cards, place the cards in a box and, while blindfolded, draw a card. Alternatively, use a table of preselected random numbers, such as Table T-12, “Two thousand random digits” of the appendix.

However, either of these methods of obtaining random numbers may be slow and cumbersome, especially if more than one random number is required. The selection process can be expedited by resorting to computer routines. Excel, for example, offers the function =RANDBETWEEN(a, b) to select a single value between, and including, a and b.

Many hand-held calculators provide a random number routine that, by the push of a button, yields a random number between 0 and 1. This function is typically labeled as “Rand” or “Rnd.” To obtain a random integer between 1 and 1000 (say), multiply the calculator’s random number by 1000, drop the fraction after the decimal point, and add 1. The reason for adding 1 is to avoid 0 (which is not on the list) and to allow the possibility of selecting 1000.

8.3 Sampling for attributes

Inference based on data measured on an interval or a ratio scale typically focuses on parameters such as the mean and the variance. Inference based on discrete data typically focuses on the composition of a population (e.g., the proportion of the population having a specific characteristic or attribute).

<i>attribute</i>	An <i>attribute</i> is a qualitative characteristic of an item that it either possesses or does not possess. This implies that every item in the population can be classified unambiguously into one of two groups. Such a classification is called <i>dichotomous</i> or <i>binary</i> . Binary classifications are modeled by the Bernoulli distribution, introduced in Section 8.5.
<i>dichotomous binary</i>	
<i>sampling for attributes</i>	<i>Sampling for attributes</i> is the process by which information about the composition of a population is investigated by classifying selected items (usually chosen at random) with a binary classification.

In regulatory work, sampling for attributes is very important: An item, record, or unit is either in compliance or not in compliance; a circuit may be open or not; a security guard may be present on location or not. Other examples of binary classification are {yes, no}, {on, off}, {defective, not defective}, {employed, unemployed}, and {married male and employed, otherwise}.

If an outcome or observation results in a grey area (e.g., maybe, almost, or undetermined), then we must modify our approach to binary classification. For example, we may classify a grey area outcome such as “slightly defective” consistently and unambiguously as one attribute or the other. In either case, it is imperative that the disposition of the “close-call” outcomes be determined before data collection in order to avoid the introduction of bias.

8.4 Sampling with and without replacement

Suppose we wish to infer the composition of a population of dichotomous items based on a random sample of n items from the population. For our statistical analysis, we need to take account of the probabilities of drawing the possible random samples. These probabilities depend on the method of drawing the sample as well as on the composition of the population. This is illustrated by a simple example. Suppose we have an urn with four marbles, any of which may be white or red. We will draw a sample of two marbles by first drawing one marble, putting it aside, and then drawing a second marble, all while we are blindfolded. Suppose further that the urn contains two white and two red marbles and that we wish to calculate the probability that the sample will consist of two white marbles. Clearly, the probability of drawing a white marble from the urn in the first draw is $1/2$. The probability of drawing a white marble in the second draw depends on the composition of the urn after the first draw. Because one white and two red marbles remain after the first draw, the probability of drawing a white marble in the second draw is $1/3$. Hence, the probability of drawing two white marbles is $(1/2)(1/3) = 1/6$.

<i>sampling without</i>	A sampling scheme where an item is not returned to its population after being drawn and classified is called <i>sampling without replacement</i> . Destructive assays are a good example
-------------------------	--

<i>replacement</i>	of sampling without replacement; a fire cracker tested for usability cannot be tested a second time. The statistical properties of sampling without replacement can be modeled by the <i>hypergeometric distribution</i> , introduced in Section 8.6.
<i>hypergeometric distribution</i>	The hypergeometric distribution and its applications are described in greater detail in Chapter 21.
<i>sampling with replacement</i>	An alternative sampling scheme is called <i>sampling with replacement</i> , where each item in the sample is returned to the population after it is drawn and classified. In the urn example above, the probability of drawing a second white marble remains unchanged so that the probability of drawing two white marbles is $(1/2)(1/2) = 1/4$. Sampling with replacement is popular in fish and wildlife migration studies, where captured animals are measured and released back to the wild, where they may be captured again. Sampling with replacement is also used in traffic control, where, after being caught, a violator may resume his trip only to be caught speeding again later. The statistical properties of sampling with replacement can be modeled by the <i>binomial distribution</i> , introduced in Section 8.7. This distribution and its applications are described in greater detail in Chapter 22. Alternatively, when the presence of an attribute is a rare event, the applicable distribution is the Poisson distribution. The Poisson distribution is introduced in Section 8.10 and described in greater detail in Chapter 23.
<i>binomial distribution</i>	

- To test your understanding, verify that if an urn contains two white and three red marbles, then:

...if we draw one marble, the probability of drawing red is $3/5$.

...if we draw (and keep) a red marble on a first trial, the probability of drawing a red marble on a second trial is $1/2$.

...if we draw (and keep) a white marble on a first trial, the probability of drawing a red marble on a second trial is $3/4$

...if each drawn marble is returned to the urn, the probability of drawing a red marble on the next draw is always $3/5$.

8.5 Bernoulli distribution

<i>Bernoulli distribution</i>	The simplest discrete distribution is the <i>Bernoulli distribution</i> , which assumes only two possible values, 0 and 1. It is used to model an experiment with two possible outcomes (often labeled
-------------------------------	--

Bernoulli trial “success” or “failure”) or an observation that either has a specified attribute or does not. The experiment or observation is sometimes called a **Bernoulli trial** or simply a **trial**. These applications are often combined by labeling the experimental outcome a “success” when the attribute is present or a “failure” when it is absent. The usual convention for the Bernoulli distribution is to label a success as “1” and a failure as “0.”

- The Bernoulli distribution is named after the Swiss theologian-cum-mathematician Jacob (a.k.a. Jakob or James or Jacques) Bernoulli, (1654–1705). He was in a family of several generations of mathematicians and physical scientists, Bernoulli is credited with coining the term *integral*.

The following are examples of Bernoulli trials:

- a coin toss that may be either “heads” or “tails”
- a statement that is either “correct” or “incorrect”
- a die roll that may be either “6” or “not 6”
- a pipe weld that is either “acceptable” or “not acceptable”
- a person who is either “infected” or “not infected”
- an accident that was either “avoidable” or “unavoidable”
- If “heads” is a success, a fair coin has probability 0.5 of success.
- If “6” is a success, a fair die has probability 1/6 of success.

If Y has a Bernoulli distribution, we use π (Greek letter pi) to denote $Pr\{1\}$. The *pf* for the Bernoulli distribution is:

$$\begin{aligned} f(1) &= Pr\{1\} = \pi \\ f(0) &= Pr\{0\} = 1 - \pi \end{aligned} \quad (8.10)$$

The mean, variance, and standard deviation of Y are:

$$\mu = \pi \quad (8.11)$$

$$\sigma^2 = \pi(1 - \pi) \quad (8.12)$$

$$\sigma = \sqrt{\pi(1 - \pi)} \quad (8.13)$$

The *pf* for a Bernoulli distribution is graphed in Figure 8.2 for $p = 0.5$ (as in a toss of a fair coin) and in Figure 8.3 for $\pi = 0.1333$ (as in a roll of a fair die if success is defined as rolling a “6”).

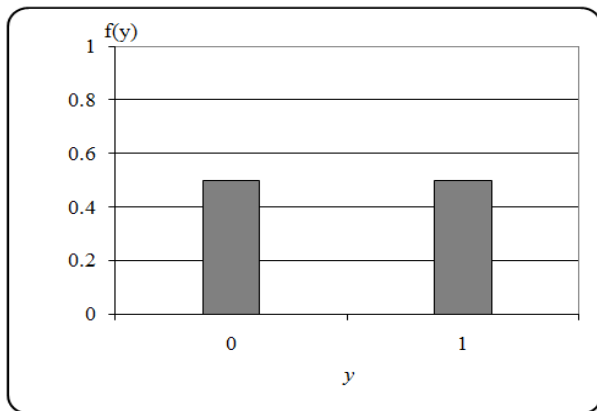


Figure 8.2. Bernoulli pf : fair coin

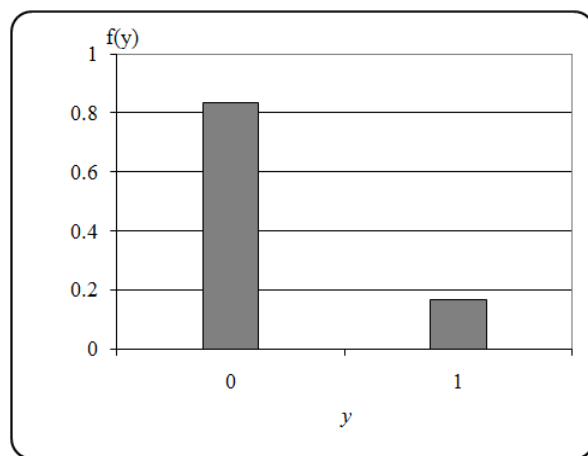


Figure 8.3. Bernoulli pf : fair die

8.6 Hypergeometric distribution

hypergeometric experiment A *hypergeometric experiment* is a sampling scheme that samples a population for attributes without replacement and which satisfies the following conditions:

- The sampled population is finite.
- Once an item is selected, it cannot be selected again.
- The size of the population is known.
- The number of items with the attribute of interest is known.
- Each item in the sample is drawn at random.

hypergeometric distribution Let Y = the number of items in the sample with the attribute of interest. Then Y has a **hypergeometric distribution**. The probability that $Y = y$, given a sample of size n and a population of N items, M of which have the attribute of interest, is:

$$f(y) = f(y; N, M, n)$$

$$= Pr\{y/N, M, n\} = \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}, \quad (8.14)$$

$$y = \max(0, n + M - N), \dots, \min(M, n)$$

- The lower limit $\max(0, n - N + M)$ placed on y in Equation (8.14) is needed if the sample size n is larger than the number of items without the attribute in the population ($N - M$). In such a case, the number of items in the sample with the attribute y is minimized if all of the items without the attribute are in the sample (i.e., $y \geq n - (N - M) > 0$). The upper limit $\min(M, n)$ placed on y in Equation (8.14) is needed because the number of items with the attribute in the sample (y) cannot be larger than the number of items with the attribute in the population (M), nor can it be larger than the number of items in the sample (n).

The mathematical expression $\binom{a}{b}$ in Equation (8.14) is called a **binomial coefficient** and is read as “ a items taken b at a time.” This expression counts the number of ways one may select b items out of a collection of a distinct items. For integers a and b with $0 \leq b \leq a$, a formula for $\binom{a}{b}$ is:

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} \quad (8.15)$$

- Note that $a!$ is the product $(1)(2) \dots (a)$ of the first a positive integers.
- Note also that $0!$ is defined as 1 (unity).
- The name “binomial coefficient” is derived from the fact that it is the coefficient of x^b in the expansion of the binomial expression $(1 + x)^a$.

The mean, variance, and standard deviation of the hypergeometric distribution are given by:

$$\mu = \frac{nM}{N} \quad (8.16)$$

$$\sigma^2 = \frac{nM}{N} \frac{N-M}{N} \frac{N-n}{N-1} \quad (8.17)$$

$$\sigma = \sqrt{\frac{nM}{N} \frac{N-M}{N} \frac{N-n}{N-1}} \quad (8.18)$$

The shape of the hypergeometric distribution depends on the parameters of the distribution: the population size (N), the number of items with the attribute in the population (M), and the sample size (n). Figures 8.4 and 8.5 show two hypergeometric distributions with the same N and M but with different sample sizes.

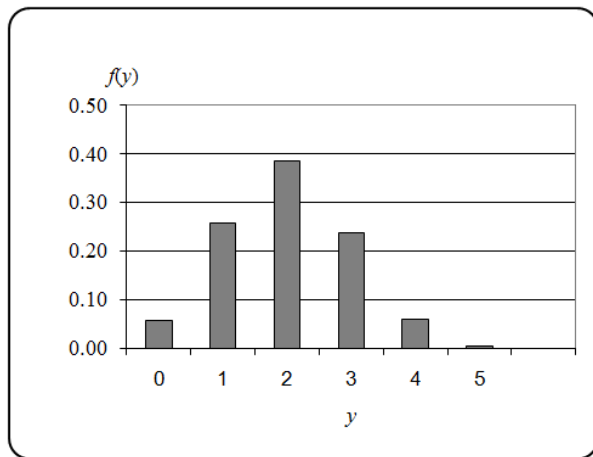


Figure 8.4. Hypergeometric pf , $N = 25$, $M = 5$, $n = 10$

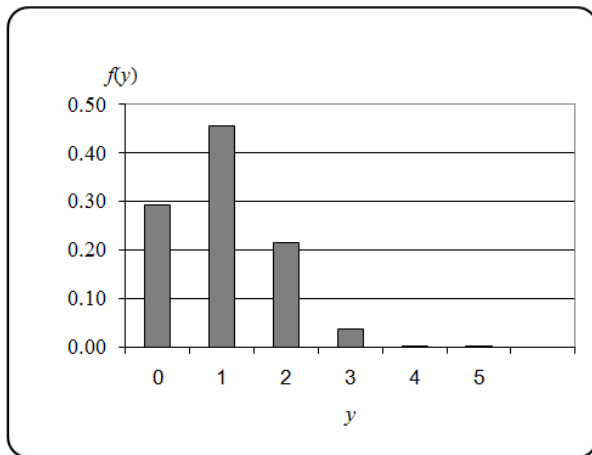


Figure 8.5. Hypergeometric pf , $N = 25$, $M = 5$, $n = 5$

The calculations for Figures 8.4 and 8.5 are somewhat lengthy. Instead, you may use Excel's function =HYPGEOMDIST(y , n , M , N) to calculate $f(y)$. The calculation used for Figures 8.4 and 8.5 are shown in Table 8.1.

Table 8.1. Hypergeometric distribution calculations

N	M	n	y	$f(y)$	N	M	n	y	$f(y)$
25	5	10	0	0.06	25	5	5	0	0.29
25	5	10	1	0.26	25	5	5	1	0.46
25	5	10	2	0.39	25	5	5	2	0.21
25	5	10	3	0.24	25	5	5	3	0.04
25	5	10	4	0.06	25	5	5	4	0.00
25	5	10	5	0.00	25	5	5	5	0.00

➤ Note that, apart from rounding error, the $f(y)$ values add up to 1.

Chapter 21 gives an extensive treatment of the hypergeometric distribution with examples and references.

8.7 Binomial distribution

binomial experiment

Let Y = the number of successes in n independent Bernoulli trials, where the probability of success, π , is constant across the n trials. This is called a *binomial experiment* and Y has a binomial distribution, denoted by $B(n, \pi)$. A binomial experiment models quality control, survey sampling, and reliability issues. The requirements for a binomial experiment are listed below, with the understanding that some of these requirements may overlap.

- The number of trials, n , is fixed.
- A binomial experiment does not allow for a variable number of trials. In contrast, some experiments continue until a fixed number of successes is achieved (see Sections 8.8 and 8.9).
- All trials are conducted in an identical manner.
- This requirement need not be taken literally. In a coin tossing experiment, for instance, we need not insist that the same coin be tossed n times. Alternatively, we could toss n identical coins once each. The important principle is that the conditions that affect the experiment's results must be the same.
- The i^{th} trial must result in either a success (recorded as $y_i = 1$) or failure (recorded as $y_i = 0$).
- This requirement means that the result of each trial must be binary, or dichotomous. Whether you toss a coin or inspect a valve for compliance, the trial must lead to a success or a failure. A coin that lands on its edge does not meet this requirement nor does a valve that is "almost completely" closed. To satisfy the requirement, it is necessary to either disregard the result (e.g., the coin) or redefine what "closed" means (e.g., the valve).
- The probability of success is constant from trial to trial.
- This requirement is not satisfied if the experimental conditions of the experiment change, e.g., if the quality of an inspection improves due to experience or deteriorates due to fatigue.
- The n trials are independent.

- The sequence of trials has no “memory.” If you win eight times in a row in a coin-tossing contest, the probability of winning the next toss is the same as for any previous toss.

The *pf* of $B(n, \pi)$ is:

$$\begin{aligned} f(y) &= f(y; n, \pi) = \Pr\{Y = y \mid n, \pi\} = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \\ &= \frac{n!}{y!(n-y)!} \pi^y (1 - \pi)^{n-y}, \quad 0 \leq \pi \leq 1, \quad y = 0, 1, \dots, n \end{aligned} \quad (8.19)$$

where:

π = probability of success in a single trial, $0 \leq \pi \leq 1$

n = number of trials

y = number of successes

The mean, variance, and standard deviation of $B(n, \pi)$ are:

$$\mu = n\pi \quad (8.20)$$

$$\sigma^2 = n\pi(1 - \pi) \quad (8.21)$$

$$\sigma = \sqrt{n\pi(1 - \pi)} \quad (8.22)$$

Two plots of the binomial distribution are shown below, both with the same sample size, $n = 10$, but with different values of π . The purpose of these figures is to show how the binomial distribution becomes more like the normal distribution as π gets closer to 0.5. In Chapter 22, we will show that the binomial approaches a normal as n increases.

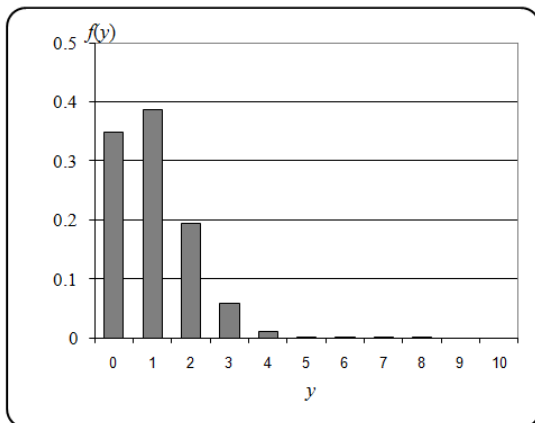


Figure 8.6. Binomial *pf*: $\pi = 0.1, n = 10$

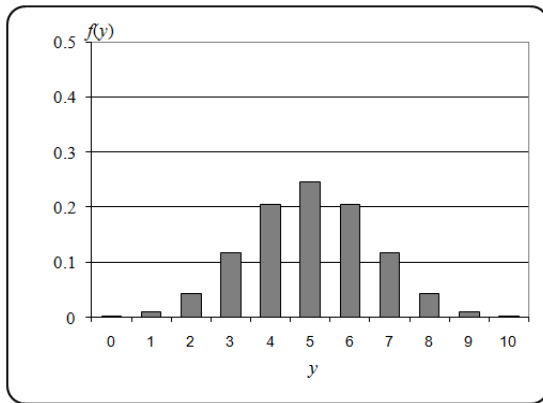


Figure 8.7. Binomial pmf: $\pi = 0.5, n = 10$

Hand calculations for Figures 8.6 and 8.7 are relatively easy, but they are even easier using Excel's function =BINOM($y, n, \pi, 0$) to calculate $f(y)$. Table 8.2 shows the probabilities that went into Figures 8.6 and 8.7. Note that the last function entry of "0" in =BINOM($y, n, \pi, 0$) is an indicator that calls for the marginal probability. If the indicator were "1," the cumulative probability would be returned.

Table 8.2. Binomial distribution probabilities

π	n	y	$f(y)$	π	N	y	$f(y)$
0.1	10	0	0.35	0.5	5	0	0.00
0.1	10	1	0.39	0.5	5	1	0.01
0.1	10	2	0.19	0.5	5	2	0.04
0.1	10	3	0.06	0.5	5	3	0.12
0.1	10	4	0.01	0.5	5	4	0.21
0.1	10	5	0.00	0.5	5	5	0.25
0.1	10	6	0.00	0.5	5	6	0.21
0.1	10	7	0.00	0.5	5	7	0.12
0.1	10	8	0.00	0.5	5	8	0.04
0.1	10	9	0.00	0.5	5	9	0.01
0.1	10	10	0.00	0.5	5	1	0.00

Chapter 22 presents an extensive treatment of the binomial distribution with examples.

8.8 Geometric distribution

geometric distribution

In a variation of a binomial experiment, continue Bernoulli trials until a success occurs. Let Y = the number of independent Bernoulli trials until the first success, with constant probability of success π . Then Y has a **geometric distribution**, denoted by $G(\pi)$. The probability that $Y = y$ is the probability that $(y - 1)$ successes are followed by a success. Accordingly, the *pf* of $G(\pi)$ is:

$$f(y) = f(y; \pi) = \Pr\{Y = y \mid \pi\} = \pi (1 - \pi)^{y-1}, \quad y = 1, 2, \dots \quad (8.23)$$

The mean, variance, and standard deviation of the geometric distribution are:

$$\mu = 1/\pi \quad (8.24)$$

$$\sigma^2 = (1 - \pi) / \pi^2 \quad (8.25)$$

$$\sigma = \sqrt{(1 - \pi) / \pi} \quad (8.26)$$

Two probability functions are graphed below, one for $\pi = 0.2$ and one for $\pi = 0.4$.

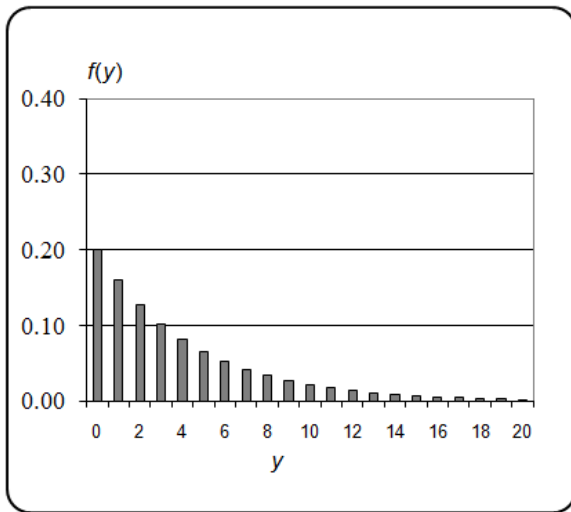


Figure 8.8. Geometric *pf*, $\pi = 0.2$

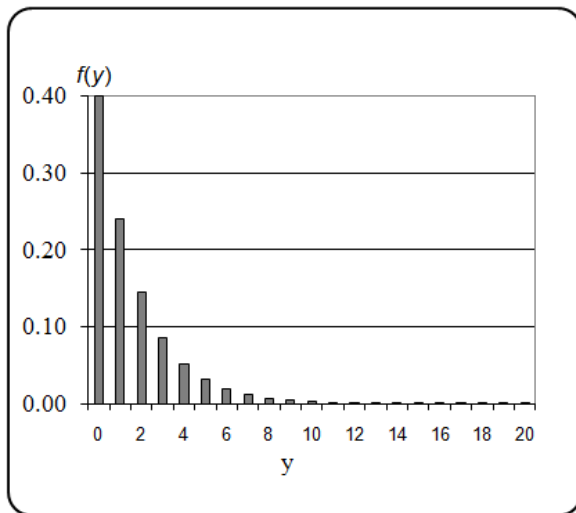


Figure 8.9. Geometric pmf: $\pi = 0.4$

Excel does not have a routine for calculating geometric distribution probabilities. However, since the geometric distribution is a special case of the negative binomial distribution (see Section 8.9), we can calculate such probabilities using Excel's =NEGBINOM($y-1, 1, \pi$) function. In that function, the middle term, 1, indicates that only one success is required after $y - 1$ failures. For example, the probability of the first success at the sixth trial when $\pi = 0.4$ is =NEGBINOMDIST(5, 1, 0.4) = 0.031. Excel's calculations for Figures 8.8 and 8.9 are summarized in Table 8.3, rounded to two decimal places.

Table 8.3. Geometric distribution probabilities

π	y	$f(y)$	π	y	$f(y)$
0.2	1	0.20	0.4	1	0.40
0.2	2	0.16	0.4	2	0.24
0.2	3	0.13	0.4	3	0.14
0.2	4	0.10	0.4	4	0.09
0.2	5	0.08	0.4	5	0.05
0.2	6	0.07	0.4	6	0.03
0.2	7	0.05	0.4	7	0.02
0.2	8	0.04	0.4	8	0.01
0.2	9	0.03	0.4	9	0.01
0.2	10	0.03	0.4	10	0.00
0.2	11	0.02	0.4	11	0.00
0.2	12	0.02	0.4	12	0.00
0.2	13	0.01	0.4	13	0.00

Example 8.3. Russian roulette. Suppose that a revolver with five chambers is loaded with a single bullet. If the chambers are spun before the trigger is pulled, the number of tries until the revolver discharges has a geometric distribution with $\pi = 0.2$. From Table 8.3, the probability that the revolver discharges on the first, second, or third try is 0.20, 0.16, or 0.13, respectively.

For further reading on the geometric distribution, see Evans, *et al.* (2000), p. 106.

8.9 Negative binomial distribution

*negative
binomial
experiment*

*negative
binomial
distribution*

Another variation of a binomial experiment is to perform Bernoulli trials until s successes occur, where s is specified before the trials begin. Let Y = the number of independent Bernoulli trials until s successes occur, where each trial has a constant probability of success π . This is called a **negative binomial experiment**, and Y has a **negative binomial distribution**, denoted by $NEGB(s, \pi)$. In contrast to a binomial experiment where the number of trials, n , is fixed and the number of successes is a random variable, a negative binomial experiment is one where the number of successes, s , is fixed and the number of trials is a random variable. One example is to play a casino game until we win three times; another is to examine oranges in a produce department until we find five ripe oranges.

- Note that the geometric distribution is a special case of the negative binomial, with $s = 1$.

The probability that $Y = y$ is the probability that there are $(s-1)$ successes in $(y-1)$ trials followed by a success in the y^{th} trial. Accordingly, the *pf* is:

$$\begin{aligned} f(y) = f(y | s, \pi) &= Pr\{y | s, \pi\} = \binom{y-1}{s-1} \pi^s (1-\pi)^{y-s} \\ &= \frac{(y-1)!}{(s-1)!(y-s)!} \pi^s (1-\pi)^{y-s}, \quad y = s, s+1, \dots \end{aligned} \tag{8.27}$$

where $y = s, s + 1, \dots$

The mean, variance, and standard deviation of the negative binomial distribution are:

$$\mu = \frac{s}{\pi} \quad (8.28)$$

$$\sigma^2 = \frac{s(1-\pi)}{\pi^2} \quad (8.29)$$

$$\sigma = \sqrt{\frac{s(1-\pi)}{\pi^2}} \quad (8.30)$$

Example 8.4. Coin tossing. Let Y be the number of tosses before three heads are obtained when tossing a fair coin. Figure 8.10 plots the *pf* of Y .

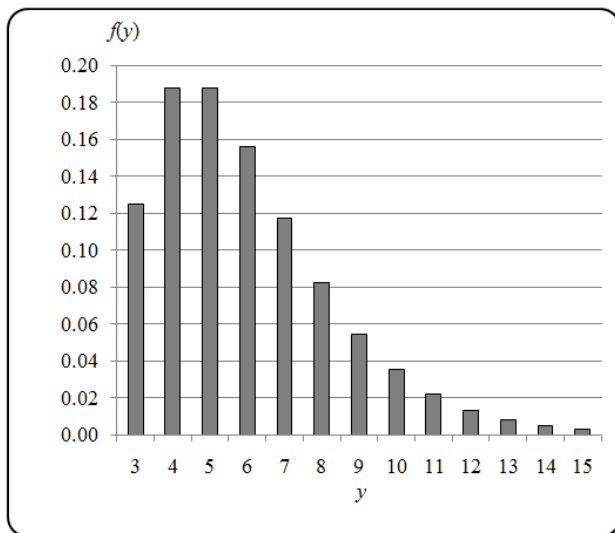


Figure 8.10. Negative binomial *pf*: $\pi = 0.5, s = 3$

Negative binomial probabilities can be calculated using Excel's =NEGBINOMDIST($y-s, s, \pi$), where s is the number of successes required, $(y - s)$ is the number of failures before s successes are reached, and π is the probability of a single success. Note that y cannot be smaller than s , as we cannot reach s successes in fewer than s trials.

Based on the use of Excel, the values used to construct Figure 8.10 are given in Table 8.4.

Table 8.4. Negative binomial probabilities

π	s	$y - s$	y	$f(y)$
0.5	3	0	3	0.125
0.5	3	1	4	0.188
0.5	3	2	5	0.188
0.5	3	3	6	0.156
0.5	3	4	7	0.117
0.5	3	5	8	0.082
0.5	3	6	9	0.055
0.5	3	7	10	0.035
0.5	3	8	11	0.022
0.5	3	9	12	0.013
0.5	3	10	13	0.008
0.5	3	11	14	0.004
0.5	3	12	15	0.003

Further discussion of the negative binomial distribution may be found in:

Krishnamoorthy (2006), p. 97

Patel et al. (1976), p. 204

Evans et al. (2000), p. 140

8.10 Poisson distribution

Poisson process A *Poisson process* models counts of events occurring in time or space. Examples are counts of system or component failures and defects on a surface or within a volume of material. Let Y = the total number of events occurring in some interval of time t with an occurrence rate θ per unit time. Instead of time, the events might be occurring in space, so that time might be replaced by length, area, or volume. The variable Y has a *Poisson distribution* with parameter $\lambda = \theta t$.

A Poisson process has several prerequisites, or assumptions, which are described in detail in Chapter 23. These assumptions can be informally stated as follows:

- The probability of an event occurring in a small interval is proportional to the length of the interval.
- The occurrence of an event in one interval is independent of that of any other disjoint interval.
- The probability of occurrence of more than one event in a small interval is negligible.

Under these assumptions, the *pf* of Y is:

$$f(y) = f(y; \lambda) = Pr\{Y = y/\lambda\} = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots \quad (8.31)$$

A remarkable property of the Poisson distribution is that its mean and variance are the same. We have:

$$E[Y] = \lambda \quad (8.32)$$

$$V[Y] = \lambda \quad (8.33)$$

Figures 8.11 and 8.12 show two Poisson probability functions, with means 1.0 and 5.0, respectively.

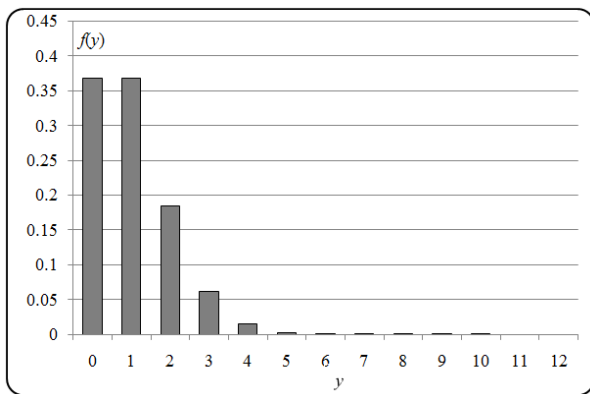


Figure 8.11. Poisson $pf, \lambda = 1$

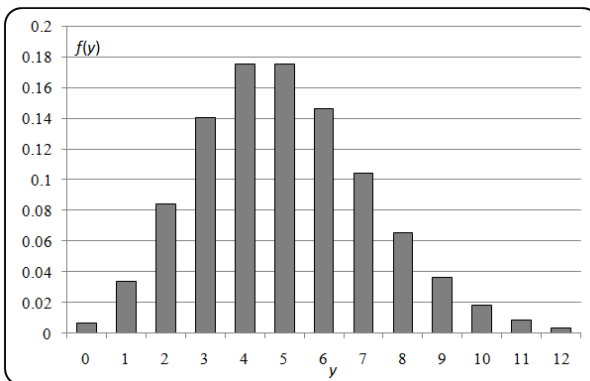


Figure 8.12. Poisson $pf, \lambda = 5$

The Poisson distribution is related to the exponential distribution introduced in Section 7.13. For the time-based Poisson defined in here, both distributions model a failure process over time. As an example, consider a component that can fail from time to time and is then repaired and returned to service. (The repair time is short enough to be considered instantaneous.) Assume the failure process satisfies the assumptions listed at the beginning of this section.

The Poisson distribution counts the number of failures in a specified time and the exponential distribution measures the time to the first failure. Under the Poisson assumptions, the time to the first failure has an exponential distribution with a failure frequency or rate that is the same as the constant failure rate of the Poisson process. This is a consequence of a key property of a Poisson process, namely, that it has no memory. This means that if the component is still functioning at any time, it remains as good as new, and its remaining life has the same distribution as when it started functioning. Once the first failure occurs, the clock starts again, and the time to the second failure has an exponential distribution that is independent of the first exponential distribution, and so on.

The relation between the Poisson and exponential distributions can therefore be stated as follows: For any fixed time period t and constant failure rate θ , let Y be the number of failures in time t . Let t_1, t_2, \dots, t_Y be the times between successive failures. Then Y has a Poisson distribution with parameter $\lambda = \theta t$ and t_1, t_2, \dots, t_Y have independent exponential distributions with parameter θ .

References to the Poisson distribution may be found in:

Krishnamoorthy (2006), p. 71

Hastings and Peacock (1975), p. 108

Patel et al. (1976), p. 21

Chapter 23 presents an extensive treatment of the Poisson distribution with examples.

9

Estimation

9.1 What to look for in the remainder of this book

Chapters 1–8 focus on distributions with known parameters, on the properties of the distributions, and on associated probability calculations. These chapters lay the groundwork for the application of statistical techniques to data, which is the subject of the remainder of this book.

Recall the example presented in Section 1.3, where probability and statistics are contrasted. In that example, while blindfolded, we draw a handful of marbles from a jar whose contents we know and make probability statements about the contents of our hand. In statistics, the process is reversed: We draw a handful of marbles from a jar, examine them, and make inferences about the contents of the jar. Starting with Chapter 9, we dip into the jar and make inferences about what is in that jar.

The jar represents a population and the contents of our hand represent data. Data are generated by experiments, processes, or observations. Here, data will be regarded as a sample from a population either with an unknown probability distribution or with an assumed probability distribution but with unknown parameters. We will use statistical inference from the data to identify relationships that help us understand and improve processes, to identify

plausible families of distributions, and to estimate the parameters of those distributions that are relevant to the issues that motivated our investigation.

Statistical inference will be used to answer questions such as:

- Are the data consistent with the requirement that a particular reactor component be able to operate for at least t hours with 99% reliability?
- Given the data we have (and the way we obtained them), how high might the probability of failure-to-start for the diesel generator reasonably be, and is that level acceptable?
- Are the apparent differences between components provided by different suppliers real, or could they just be random?

9.2 What to look for in Chapter 9

In this chapter, we are first reminded of several terms that are already in our vocabulary:

- *parameter*, §9.4
- *estimator*, §9.4
- *estimate*, §9.4

Some properties of estimators are then explained:

- *point estimator*, §9.5
- *unbiased estimator*, §9.5
- *minimum variance estimator*, §9.5
- *interval estimator*, §9.6

Some practical concepts and procedures will be defined and illustrated:

- *estimation and inference*, §9.3
- *confidence intervals for a mean*, §9.6 – §9.11
- *tolerance limits for a normal distribution*, §9.12
- *confidence intervals for a variance*, §9.13
- *sample size determination*, §9.14, §9.15

9.3 Estimation and inference

estimation The principal objectives of statistics are often characterized as *estimation* and *inference*. These two objectives are strongly interrelated. In their practical aspects, they both involve the design, collection, organization, analysis, presentation, and interpretation of data. In their theoretical bases, they both depend on mathematical formulations of probabilistic and statistical concepts. Estimation

inference

and inference are distinguishable as follows:

The goal of *estimation* is to assign values to population parameters.

The goal of *inference* is to draw conclusions about the population from which the data came.

Chapter 9 focuses on estimation processes; inferential processes are described in Chapter 10. Once established, both concepts are explored in a variety of analyses in the remainder of the book.

9.4 Elements of estimation

Statistical estimation consists of three elements. These elements were introduced earlier (Section 1.6) and are now restated.

parameter A *parameter* is a quantity we wish to estimate. In Government parlance, a parameter is an “estimee.” While both populations and random variables have parameters, in this book, we estimate only population parameters. A function of a parameter is also a parameter (e.g., the variance σ^2 is a parameter and so are σ and 3σ). Random variables and their associated distributions are mathematical tools used to model samples and populations. Unless stated otherwise, “random variable” and “distribution” will be used only in the context of a discussion about a sample or a population.

Examples of parameters are a population mean μ and a population proportion π .

Specific examples of parameters are:

- Average weight μ of all ice baskets in an ice condenser reactor
- Proportion π of all employees who would test positive in a drug test if the test were conducted today

➤ Wherever practical, we use Greek letters to denote parameters.

estimator An *estimator* is a function of data (usually expressed by a formula) used to estimate a population parameter. Because data are modeled by a random variable, an estimator is also a random variable. The data come from a *random sample*, which is a set of independent observations from a specified population or the distribution that models it. (Unless stated otherwise, a random sample will sometimes be simply referred to as a *sample*.)

Examples of estimators are a sample mean \bar{Y} and a sample variance S^2 .

Specific examples of an estimator are:

- Average response time \bar{Y} to an alarm, obtained from a random sample of challenges
- Proportion P of payable invoices that were not paid within 30 days of receipt, obtained from a random sample of invoices
- Wherever practical, we use upper case Latin letters to denote estimators. However, sometimes an estimator is written as the parameter with a “hat”, such as $\hat{\pi}$. The use of a hat is appropriate as it reminds us that this is not a parameter (an unknown constant) but is rather an estimator (a random variable) used to estimate a parameter.

estimate An *estimate* is the numerical value obtained by applying an estimator to the data. Once calculated, an estimate is a numerical value (i.e., it is an observation of the random variable defined by the estimator).

Specific examples of estimates are:

- A calculated value $\bar{y} = 67.1$, the average height (in inches), based on a sample of students in the school
- A calculated proportion $p = 0.40$ of responses that were incomplete, based on a sample of examination questions
- Wherever practical, we use lower case Latin letters to denote estimates.

There are two major categories of statistical estimators: point estimators and interval estimators. These are discussed in Sections 9.5 and 9.6, respectively.

9.5 Point estimators

point estimator A *point estimator* is a rule that produces a single number to estimate a population parameter. Given a sample from a population, examples are:

- The sample mean \bar{Y} is a point estimator of the population mean.
- The sample variance S^2 is a point estimator of the population variance.
- The proportion P of defects in the sample is a point estimator of the population proportion of defects.

Books on theoretical statistics have identified a number of desirable properties of estimators. Those properties may be used to choose an estimator from among several competing estimators of the same parameter. This book, however,

focuses on only two important properties (introduced below) of an estimator: being unbiased and having minimum variance. Other properties of estimators are discussed in texts such as Wilks (1963), Chapter 12.

unbiased A point estimator is said to be **unbiased** if its expected value is equal to the parameter of interest. This means that, theoretically, the average value of estimates, based on many samples of the same size repeatedly drawn from the same population under identical conditions, is equal to the parameter of interest.

➤ Unless the sample consists of only one observation, there is almost always an infinite number of unbiased estimators of the same parameter based on the sample. In such cases, another criterion may be used to choose an unbiased estimator that is “better” than other unbiased estimators.

minimum variance A point estimator is said to have **minimum variance** if it has a variance no larger than that of other estimators of the same parameter from the same sample.

minimum variance unbiased A point estimator is said to be **minimum variance unbiased** if it is both unbiased and has minimum variance.

Some facts about estimators:

- The sample mean \bar{Y} is an unbiased estimator of the population mean μ .
 - If the random variable that models the population is symmetric about its mean, the sample midrange, mode, and median are also unbiased estimators of μ .
 - The sample mean \bar{Y} has a variance σ^2/n that is no larger than the variance of any other unbiased estimator of μ . Thus, \bar{Y} is a minimum variance unbiased estimator of μ .
 - The sample variance S^2 is an unbiased estimator of a population's variance σ^2 .
- Because square root is a nonlinear function, S is not an unbiased estimator of a population's standard deviation σ .

The following example is used here to illustrate point estimators and later in the chapter to illustrate interval estimators and tolerance limits. The principles applied here are applicable to all types of distributions, not just to the normal distribution properties.

Example 9.1. Natural uranium concentration. Natural uranium concentration, in parts per million (ppm), was measured in a random sample of

41 wells in a neighborhood. Table 9.1 reports the measurements in ascending order. We wish to estimate the mean and variance of natural uranium concentration in the neighborhood, denoted by μ and σ^2 , respectively.

Table 9.1. Natural uranium concentration (ppm)

0.012	0.012	0.074	0.090	0.099	0.103	0.109	0.117	0.127	0.139
0.148	0.156	0.159	0.165	0.189	0.197	0.200	0.209	0.220	0.223
0.250	0.255	0.261	0.268	0.283	0.286	0.292	0.293	0.309	0.310
0.320	0.346	0.366	0.402	0.425	0.436	0.440	0.611	0.661	0.689
0.750									

From the $n = 41$ observations in Table 9.1, we may obtain several point estimates of μ , such as $\bar{y} = 0.2683$, median = 0.250, and midrange = 0.3810. If the distribution from which this sample is drawn is symmetric, then all these estimators are unbiased. However, our preferred estimate of μ is $\bar{y} = 0.2683$, because it was derived from \bar{Y} , which is an unbiased minimum variance estimator of μ . An unbiased estimate of σ^2 is the sample variance $s^2 = 0.0305$.

9.6 Interval estimators

A point estimator may well be a satisfactory tool to estimate a population parameter. However, a point estimate, even if calculated from a large amount of data, is very rarely equal to the parameter being estimated. Furthermore, a point estimate may not even be a good approximation to the parameter being estimated. An alternative to a point estimator is an interval estimator that produces an interval containing the parameter of interest with a specified probability. The balance of this chapter will focus on interval estimators.

<i>interval estimator</i>	An <i>interval estimator</i> is a function of data that produces one or two finite bounds on a population parameter. The intervals can be one-sided (a lower bound or an upper bound) or two-sided (both upper and lower bounds).
<i>confidence interval</i>	The most widely used type of interval estimator is called a <i>confidence interval</i> . A confidence interval is accompanied by a probabilistic statement that provides a measure of assurance that the interval contains the parameter being estimated. This probabilistic measure of assurance is called the <i>confidence level</i> or <i>confidence coefficient</i> . The end-points of a confidence interval are called <i>confidence limits</i> or <i>confidence bounds</i> .
<i>confidence level</i>	
<i>confidence coefficient</i>	
<i>confidence limits</i>	
<i>confidence bound</i>	

Note that various writers, and we are no exception, use a variety of notations and expressions in quantifying the confidence coefficient. For instance, we see

phrases such as “95% confident,” “95% confidence,” “95% sure,” or, occasionally, “0.95 confidence.”

For a more detailed discussion of interval estimators, see Kendall and Buckland (1971), p. 145, or Hahn and Meeker (1991).

two-sided confidence interval Confidence intervals that state both finite upper and lower confidence limits are called **two-sided confidence intervals**. Each of the following statements is an example of a two-sided confidence interval:

- We are 95% confident that the mean population weight is between 9.4 and 10.1 kilograms.
- We have 90% confidence that the proportion of cables that are traceable to control unit X is between 7% and 11%.
- We are 0.9999 confident that the percentage of plutonium by weight in a batch of plutonium dioxide lies in the interval (87.16, 89.02).

one-sided confidence intervals Confidence intervals that state either an upper or a lower finite confidence limit are called **one-sided confidence intervals**. Each of the following statements is an example of a one-sided confidence interval:

- We are 75% confident that the mean-time-between-failure of electric motors produced by a specific manufacturer is at least 28,600 hours.
- We are 95% sure that the average cost of travel is less than \$925. (Managers may use this figure in travel budget projections.)

lower confidence bound The first of the one-sided confidence bounds above is called a **lower confidence bound** or a **lower confidence limit**. We could also write the interval as $(28,600, \infty)$, showing a finite value of 28,600 as the lower limit and an unbounded upper limit, even though we know that no motor will function forever.

upper confidence bound The second of the one-sided confidence bounds above is called an **upper confidence bound** or an **upper confidence limit**. We could also write the interval as $(-\infty, 925)$, which is mathematically correct but may raise eyebrows.

9.7 Confidence intervals for a mean

Suppose we have a random sample of size n with sample mean \bar{Y} from a population with mean μ and standard deviation σ . We wish to construct a confidence interval for μ based on \bar{Y} . However, to clarify the general construction, we will start by constructing special confidence intervals under some specific assumptions. We will then relax these assumptions one by one to construct more general confidence intervals. The assumptions we make at this point are:

Assumption 1. \bar{Y} is distributed normally.

Assumption 2. A two-sided confidence interval for μ is required.

Assumption 3. The confidence level is 95%.

Assumption 4. The standard deviation σ is known.

Assumption 1 holds whenever the population is distributed normally. However, even when the population is not distributed normally, if the sample size is “large,” the sample mean will be approximately normal, courtesy of the Central Limit Theorem (Section 7.8). In such cases, the confidence intervals based on Assumption 1 will be approximately correct. For cases where this assumption does not hold, we may be able to use a nonparametric approach (see Chapter 25).

All the confidence intervals constructed in this chapter are based on Assumption 1. We can test whether the sample came from a normal population by applying one of several available procedures, some of which are given in Chapter 11.

9.8 Two-sided 95% confidence intervals for a mean

When all four assumptions of Section 9.7 are met, we may proceed to construct a 95% two-sided confidence interval for μ . To that end, we look up Table T-1 of the appendix, where the 0.975 quantile is $z_{0.975} = 1.960$. This quantile sets off 2.5% of the distribution at the right end (sometimes called the “tail”) of the distribution. By symmetry, 2.5% of distribution lies to the left of $z_{0.025} = -1.960$, and therefore 95% of the standard normal distribution lies between -1.960 and $+1.960$.

- As stated in Chapter 7, we may use Excel’s function `=NORMSINV(q)` to obtain the desired quantile. We find that `=NORMSINV(0.975)` returns 1.960 and `=NORMSINV(0.025)` returns -1.960. Accordingly, we have the following inequality:

$$Pr\{-1.960 < Z < 1.960\} = 0.95 \quad (9.1)$$

where Z is the standard normal variable:

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \quad (9.2)$$

Applying Equation (9.2) to Equation (9.1), we have:

$$Pr\left\{-1.960 < \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} < 1.960\right\} = 0.95 \quad (9.3)$$

Algebraic manipulation of the inequality in the brackets in Equation (9.3) leads to the following successive inequalities:

$$\begin{aligned} -1.960\sigma / \sqrt{n} < \bar{Y} - \mu < 1.960\sigma / \sqrt{n} \\ 1.960\sigma / \sqrt{n} > \mu - \bar{Y} > -1.960\sigma / \sqrt{n} \\ \bar{Y} + 1.960\sigma / \sqrt{n} > \mu > \bar{Y} - 1.960\sigma / \sqrt{n} \end{aligned}$$

One more reversal of the last inequality yields the conventional form of a confidence interval (i.e., with the lower limit at the left and the upper limit at the right):

$$\bar{Y} - 1.960\sigma / \sqrt{n} < \mu < \bar{Y} + 1.960\sigma / \sqrt{n} \quad (9.4)$$

Written in interval form, the inequality in Equation (9.4) is a two-sided 95% confidence interval for the mean μ of a normal population when the standard deviation σ is known.

$$\left(\bar{Y} - 1.960\sigma / \sqrt{n}, \bar{Y} + 1.960\sigma / \sqrt{n}\right) \quad (9.5)$$

The interval in Equation (9.5) is sometimes written as:

$$\bar{Y} \pm 1.960\sigma / \sqrt{n} \quad (9.6)$$

Note that the confidence interval is centered at the sample mean \bar{Y} . Its width (or length, if we wish) is:

$$\left(\bar{Y} + 1.960\sigma / \sqrt{n}\right) - \left(\bar{Y} - 1.960\sigma / \sqrt{n}\right) = 2(1.960\sigma) / \sqrt{n} \quad (9.7)$$

From Equation (9.7), we note that when σ is known, the width of the confidence interval is constant since it is free of random variables. Repeating the sampling

process under the same assumptions will yield a different sample mean, but the width will remain the same.

The interpretation of a 95% two-sided confidence interval is as follows: If we were to repeat the sampling process any number of times, we would construct a number of intervals, each centered on a sample mean but all having the same width as given in Equation (9.7). Because of the random nature of the sample mean, some of these intervals would include μ while others would not. From theoretical considerations, the probability that any one of these intervals would contain the parameter μ is 0.95. In this sense, we have 95% confidence that the two-sided interval we constructed includes μ .

Suppose that the variance σ^2 of natural uranium concentration of Example 9.1 is known to be $\sigma^2 = 0.05$ so that $\sigma = 0.2236$. Using the sample, $\bar{y} = 0.2683$, with $n = 41$, we calculate the end points of a 95% two-sided confidence interval as:

$$0.2683 - (1.960)(0.2236) / \sqrt{41} = 0.1999$$

$$0.2683 + (1.960)(0.2236) / \sqrt{41} = 0.3368$$

In this example, we do not use the sample standard deviation to estimate σ because σ is known.

Example 9.2. Beam momentum measurements (Frodesen et al. (1979), p. 171). Measurements of the momentum of monoenergetic beam tracks on bubble chamber pictures have led to the following sequence of 10 readings, in GeV/c : {18.87, 19.55, 19.32, 18.70, 19.41, 19.37, 18.84, 19.40, 18.78, 18.76}. Based on extensive previous experience, we assume that this sample originated from a normal population with a known standard deviation of $\sigma = 0.3$ GeV/c. Find a two-sided 95% confidence interval for the beam momentum, μ .

The calculated sample mean for $n = 10$ is $\bar{y} = 19.100$. Since σ is given, there is no need, at least not at this point, to calculate the sample standard deviation. The lower and upper 95% confidence limits for μ , respectively, are therefore calculated as:

$$19.100 - (1.960)(0.3) / \sqrt{10} = 18.914$$

$$19.100 + (1.960)(0.3) / \sqrt{10} = 19.286$$

Thus, a 95% confidence interval for μ is (18.9, 19.3), where the units are in GeV/c.

In the construction of a 95% two-sided confidence interval, we elected to split off 2.5% of the distribution of \bar{Y} at the lower end and 2.5% at the upper end of

the distribution. Using equal probability splits at the ends of the distribution is the usual way of constructing a confidence interval, and its advantage is that the resulting confidence interval is shorter for any other split, such as 4% and 1%. For example, with a 1% and 4% split, the confidence interval would be:

$$(\bar{Y} - z_{0.99}\sigma/\sqrt{n}, \bar{Y} + z_{0.99}\sigma/\sqrt{n}) = (\bar{Y} - 2.32\sigma/\sqrt{n}, \bar{Y} + 2.32\sigma/\sqrt{n})$$

with width $4.077\sigma/\sqrt{n}$. From Equation (9.7), this is larger than the width of $3.920\sigma/\sqrt{n}$ of the 95% confidence interval with an even split.

9.9 One-sided 95% confidence intervals for a mean

Two-sided confidence intervals (Assumption 2) are not always appropriate. Many, if not most, practical applications require a one-sided confidence interval to guard against excess in one direction (either too much or too little). We construct a 95% one-sided confidence interval for μ based on the assumptions of normality (Assumption 1), 95% confidence requirement (Assumption 3), and knowledge of σ (Assumption 4).

A 95% lower confidence limit for μ is:

$$\bar{Y} - 1.645\sigma/\sqrt{n} \tag{9.8}$$

A 95% upper confidence limit for μ is:

$$\bar{Y} + 1.645\sigma/\sqrt{n} \tag{9.9}$$

where the number 1.645 is the 0.95 quantile of the standard normal distribution (Table T-1) or from Excel's =NORMSINV(0.95).

Applying Equation (9.8) to the data in Example 9.1 where $\bar{y} = 0.2683$ and the known $\sigma = 0.2236$, yields the lower 95% confidence limit for μ , and Equation (9.9) yields the upper 95% confidence limit:

$$0.2683 - (1.645)(0.2236)/\sqrt{41} = 0.2109$$

$$0.2683 + (1.645)(0.2236)/\sqrt{41} = 0.3258$$

Similarly, for Example 9.2, where $\bar{y} = 19.100$ and $\sigma = 0.3$, we obtain the lower and upper 95% one-sided confidence limits:

$$19.100 - (1.645)(0.3)/\sqrt{10} = 18.944$$

$$19.100 + (1.645)(0.3)/\sqrt{10} = 19.256$$

The interpretation of a one-sided 95% confidence interval is similar to that of the two-sided case: If we were to repeat the sampling process m times, we would construct m intervals. Because of the random nature of the data, some of these intervals would contain μ while others would not. From theoretical considerations, the probability that any one of these intervals would contain the parameter μ is 0.95. This means that if m is large, about 95% of the intervals would contain μ .

9.10 Confidence intervals with an arbitrary confidence level

The 95% level is the most common confidence level used by the NRC and as such is regarded as a default value. However, the confidence level can be any value between 0 and 1 (0% and 100%), and the selection of that value should be made considering management priorities and constraints. Thus, if our two-sided confidence quantile is $100q\%$, you can consult a standard normal table to find $z_{1-(1-q)/2} = z_{(1+q)/2}$, or obtain this value from Excel's `=NORMSINV((1+q)/2)` function. The confidence limits of the two-sided confidence interval for μ are then calculated as:

$$\bar{Y} \pm z_{(1+q)/2} \sigma / \sqrt{n} \quad (9.10)$$

In Example 9.1, suppose a two-sided 99% confidence interval is desired. We look in Table T-1 and find $z_{0.995} = 2.576$, and then calculate:

$$0.2683 \pm (2.576)(0.2236) / \sqrt{41}$$

yielding the 99% two-sided confidence interval for μ of (0.1783, 0.3583).

In Example 9.2, suppose a two-sided 98% confidence interval is desired. We find $z_{0.99} = 2.326$ and calculate:

$$19.100 \pm (2.326)(0.3) / \sqrt{10}$$

yielding the 98% two-sided confidence interval for μ of (18.878, 19.321).

For a 90% one-sided confidence limit for μ , the associated z value is $z_{0.90} = 1.282$. The lower confidence limit for μ in Example 9.2 is calculated as $19.100 - (1.282)(0.3) / \sqrt{10} = 18.977$. Similarly, the 90% upper confidence limit for μ is $19.100 + (1.282)(0.3) / \sqrt{10} = 19.222$.

It is often convenient to refer to the complement of q , denoted by α , in looking up quantiles in a table. For that matter, we may find texts where some table headings are labeled quantile (q), some are labeled α , while still others show both labels.

9.11 Confidence intervals for unknown σ

Assumption 4 of Section 9.7 states that the standard deviation σ is known. This assumption is rarely satisfied, and hence the equations used in Sections 9.8–9.10 for constructing confidence intervals are rarely applicable. Lacking knowledge of σ , we modify the Z -statistic by replacing σ in those equations by the sample standard deviation S , given by Equation (2.12) as:

$$S = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n-1}} = \sqrt{\frac{\sum Y^2 - (\sum Y)^2 / n}{n-1}} = \sqrt{\frac{n \sum Y^2 - (\sum Y)^2}{n(n-1)}} \quad (9.11)$$

where S has $\nu = (n - 1)$ degrees of freedom.

Equation (9.2) is modified accordingly. Instead of Z , the modified standard statistic is called a T -statistic and is written as:

$$T = \frac{\bar{Y} - \mu}{S / \sqrt{n}} \quad (9.12)$$

Where T is sometimes written as T_ν to remind us of the associated degrees of freedom.

The distribution of the T -statistic is given in Section 7.11. Quantiles of the T -statistic are obtained from Table T-3 of the appendix. As an example, the quantile that corresponds to a 95% two-sided confidence interval when $\nu = 9$ is $t_{0.975}(9) = 2.26$. Excel's =TINV function returns the complementary quantile and always assumes that a two-sided interval is requested. Hence, to get $t_q(\nu)$, the call to Excel is =TINV(2(1 - q), ν). Thus, =TINV(0.050, 9) = 2.26.

In parallel with the construction of confidence intervals when σ is known, we have the following confidence intervals when σ is not known and is estimated by S .

For a 100 q % two-sided confidence interval for μ :

$$(\bar{Y} - t_{(1+q)/2}(\nu) s / \sqrt{n}, \bar{Y} + t_{(1+q)/2}(\nu) s / \sqrt{n}) \quad (9.13)$$

For a 100 q % lower confidence limit for μ :

$$\bar{Y} - t_q(\nu) s / \sqrt{n} \quad (9.14)$$

For a 100 q % upper confidence limit for μ :

$$\bar{Y} + t_q(\nu) s / \sqrt{n} \quad (9.15)$$

We return to Example 9.1 and use the sample standard deviation s to estimate σ . In Example 9.1, $\bar{y} = 0.2683$ and $s = 0.1747$, with $\nu = n - 1 = 40$ degrees of freedom. We construct 95% two-sided and one-sided confidence intervals for μ . The required quantiles are obtained from Table T-3 as $t_{0.975}(40) = 2.02$ and $t_{0.95}(40) = 1.68$ or from Excel's =TINV function, as =TINV(0.05, 40) = 2.02 and =TINV(0.1, 40) = 1.68.

For a 95% two-sided confidence interval for μ , Equation (9.13) yields:

$$0.2683 \pm (2.02)(0.1747) / \sqrt{41} = (0.2132, 0.3234)$$

For a 95% lower confidence limit for μ , Equation (9.14) yields:

$$0.2683 - (1.68)(0.1747) / \sqrt{41} = 0.2224$$

For a 95% upper confidence limit for μ , Equation (9.15) yields:

$$0.2683 + (1.68)(0.1747) / \sqrt{41} = 0.3142$$

Example 9.3. Weight of ice baskets. Out of more than 1,000 ice baskets that surround an ice condenser reactor, 25 ice baskets were selected at random to estimate the average basket weight. Construct a lower 95% confidence limit for the average basket weight. This limit, when multiplied by the number of baskets around the reactor, will provide 95% assurance that at least that amount of ice is available for cooling in case of emergency. Table 9.2 gives the weights, in kilograms, of the sample of 25 ice baskets. Assume that the weights of the ice baskets are normally distributed.

- To maintain confidentiality of information, all numbers in this example are greatly distorted.

Table 9.2. Weight of ice baskets (in kg)

892	893	894	900	901	901	913	914	916	924	931
932	932	932	935	936	938	941	942	943	943	946
950	951	963								

The statistics from this dataset are $\bar{y} = 926.52$, $s = 20.51$, and $n = 25$. From Table T-3, $t_{0.95}(24) = 1.71$. The 95% lower limit for μ is:

$$926.52 - (1.711)(20.51) / \sqrt{25} = 919.5$$

Thus, if N baskets surround the reactor, we are 95% sure that $919.5 N$ kg are available for cooling.

The interpretation of a $100q\%$ confidence interval when σ is not known is the same as when σ is known. If the sampling procedure were repeated m times and m confidence intervals were constructed, then the probability that any one of the m confidence intervals would contain μ is q .

9.12 Statistical tolerance limits for a normal population

tolerance limits The quality of a manufactured product is often specified by setting a range, the bounds of which are called **tolerance limits**. These limits have the property that a specified percentage of the product is expected to fall within them. For example, in an NRC application, we might require that at least 95% of a specific set of cable diameters fall between 3.52 millimeters (mm) and 3.84 mm. As another example, we might specify that the response time to a fire alarm must be less than 12 minutes at least 90% of the time.

If we have the entire population at our disposal, we can easily check whether our requirements are met. For instance, if we know the diameters of all the cables in the specific set, we can determine what percentage is within the limits. If we have a long history of fire responses and if conditions remained unchanged, then we can check if the 90% criterion is met. However, in practice, we may have only a sample from the population on which to base our evaluation.

statistical tolerance limits In such cases, we can construct **statistical tolerance limits** and a **tolerance interval**. Given a random sample from a population, a tolerance interval (L, U) is such that we have $100\gamma\%$ confidence that (L, U) contains at least $100\pi\%$ of the population, where $L, U, \gamma,$ and π are specified. Such an interval is often denoted as a $100\gamma/100\pi$ tolerance interval (e.g., a 95/95 tolerance interval).

Note the difference between a statistical tolerance interval and a confidence interval. A confidence interval is used to estimate a population parameter, while a statistical tolerance interval is used to make a statement about a specified proportion of items in a population. However, both types of intervals have associated confidence levels, because both are based on estimates and are therefore uncertain.

two-sided tolerance interval A tolerance interval that has both upper and lower finite tolerance limits is called a **two-sided tolerance interval**. Each of the following statements is a two-sided tolerance interval:

- We are 95% confident that at least 95% of cables diameters are between 3.52 mm and 3.84 mm (a 95/95 tolerance interval).
- We are 90% confident that at least 99% of our employees earn between \$39,014 and \$57,755 per annum (a 90/99 tolerance interval).

- We are 99% confident that the annual per-house savings from residential energy conservation is between 156 and 1,942 kilowatt-hours for at least 85% of the population (a 99/85 tolerance interval).

one-sided tolerance interval A tolerance interval that has either an upper or a lower finite tolerance limit is called a **one-sided tolerance interval**. Each of the following statements is a one-sided tolerance interval:

- We are 75% confident that at least 90% of the response time to a fire alarm will be less than 12 minutes (a 75/90 upper tolerance limit).
- We are 95% confident that at least 80% of the crates weigh more than 1.2 tons (a 95/80 lower tolerance limit).

The 95/95 specification is the most common specification for tolerance intervals at the NRC. It is usually regarded as the default tolerance interval specification.

To calculate tolerance limits, it is necessary to specify the distribution of the population. For most applications, the population is assumed to be normally distributed. The tolerance limits are based on factors k_2 , for two-sided tolerance intervals, and k_1 , for one-sided tolerance intervals. These are called **tolerance limit factors**. The factors are functions of the sample size n , the confidence level γ , in percent, and the desired population proportion π , also in percent. The factors k_2 and k_1 are found in Tables T-11a and T-11b of the appendix, respectively.

More extensive tables of the tolerance factors may be found in Odeh and Owen (1980) on pp. 85–113 for k_2 and on pages 17–69 for k_1 . Approximations to these coefficients are given by Howe (1969), p. 610, for two-sided, and by Natrella (1963), pp. 2–15 for one-sided tolerance limits.

Tolerance limits are calculated as follows:

$\bar{Y} \pm k_2 S$	for two-sided tolerance limits, where k_2 is taken from Table T-11a
$\bar{Y} + k_1 S$	for an upper tolerance limit, where k_1 is taken from Table T-11b
$\bar{Y} - k_1 S$	for a lower tolerance limit, where k_1 is taken from Table T-11b

Suppose we have a random sample of size $n = 10$ from the normal distribution with $\bar{y} = 120.5$ and $s = 8.8$. For a 95/90 two-sided tolerance interval, we use Table T-11a under $\gamma = 0.95$, $\pi = 0.90$, and $n = 10$ to find $k_2 = 2.856$. The desired

lower and upper tolerance limits are calculated as $120.5 - (2.856)(8.8) = 95.4$ and $120.5 + (2.856)(8.8) = 145.6$, respectively.

For a 95/90 one-sided tolerance limit, we use Table T-11b under $\gamma = 0.95$, $\pi = 0.90$, and $n = 10$ to find $k_1 = 2.355$. A lower tolerance limit is calculated as $120.5 - (2.355)(8.8) = 99.8$. An upper tolerance limit is calculated as $120.5 + (2.355)(8.8) = 141.2$.

Example 9.4. Tolerance limit for the critical power ratio. Critical power can be calculated by an algorithm or measured directly. The ratio of the calculated to the measured values, called the Experimental Critical Power Ratio (ECPR), was determined from a random sample of 150 measured values with a mean of 1.0211 and a standard deviation of 0.0061. Find the 95/95 upper tolerance limit for ECPR.

Table T-11b shows the 95/95 factor for a one-sided tolerance limit for $n = 150$ to be 1.870. The upper tolerance limit for the ECPR is therefore calculated as $1.0211 + (1.870)(0.0061) = 1.033$.

9.13 Confidence intervals for a variance

We have learned thus far to put bounds on a population mean, μ . At times, however, we need to place a bound on a population variance. We need to know when the variability of a measurement system is out of control, even if the mean appears stable. To construct such bounds, we refer to some theoretical findings.

Theory shows (Mood et al. (1974), p. 245) that the ratio $(n - 1)S^2/\sigma^2$ has a chi-square distribution with $\nu = (n - 1)$ degrees of freedom. We thus have the following relation that occurs with probability $1 - \alpha$:

$$\chi_{\alpha/2}^2(\nu) < \frac{\nu S^2}{\sigma^2} < \chi_{1-\alpha/2}^2(\nu) \quad (9.16)$$

Algebraic manipulation of Equation (9.16) leads to the following:

$$\frac{1}{\chi_{1-\alpha/2}^2(\nu)} < \frac{\sigma^2}{\nu S^2} < \frac{1}{\chi_{\alpha/2}^2(\nu)} \quad (9.17)$$

Accordingly, a $100(1 - \alpha)\%$ two-sided confidence interval for σ^2 is:

$$\left(\frac{\nu S^2}{\chi_{1-\alpha/2}^2(\nu)}, \frac{\nu S^2}{\chi_{\alpha/2}^2(\nu)} \right) \quad (9.18)$$

One-sided confidence intervals are similarly obtained:

$$\frac{\nu S^2}{\chi_{1-\alpha}^2(\nu)}, \text{ for a } 100(1-\alpha)\% \text{ lower confidence limit for } \sigma^2 \quad (9.19)$$

$$\frac{\nu S^2}{\chi_{\alpha}^2(\nu)}, \text{ for an } 100(1-\alpha)\% \text{ upper confidence limit for } \sigma^2 \quad (9.20)$$

The confidence limits for σ are the square roots of the limits for σ^2 .

Almost always we need an upper rather than a lower confidence limit on σ^2 or σ . We are concerned about excessive variability rather than a process that is relatively stable. (One exception to this rule is the stock market, where some investors might want to take advantage of market volatility.)

As an example, suppose a random sample of size 10 (for which $\nu = 9$) has a standard deviation of $s = 1.47$, and we want a 95% upper confidence limit for σ . First, from Table T-2, the 0.05 quantile for the chi-square distribution with $\nu = 9$ degrees of freedom is 3.33. Next, the upper confidence limit for σ^2 is calculated as $(9)(1.47)^2/3.33 = 5.840$. Finally, the desired upper limit for σ is $\sqrt{5.840} = 2.417$.

- Note that confidence limits for σ^2 (or σ) do not depend on μ or \bar{Y} . This makes sense even though confidence limits for μ depend on σ (if σ is known) or S (if σ is not known). Confidence limits for a parameter always depend on the variability in the estimator of the parameter and thus depend on σ both for μ and σ . The mean μ is a measure of the location of the distribution and does not affect its variability.

The interpretation of a 95% confidence interval for σ^2 (or σ) is the same as for a confidence interval for μ . If we were to repeat the sampling process any number of times, each time constructing a 95% confidence interval for σ^2 (or σ), then each of these confidence intervals has probability 0.95 of containing σ^2 (or σ).

9.14 Sample size determination: σ known

Perhaps the most common question asked of a statistician is “How large a sample do I need?” The answer is not simple. It depends on several factors and assumptions. This section considers a situation in which our problem is the estimation of the mean of a population. Furthermore, we seek reasonable assurance that the resulting estimate of the mean does not deviate from the true mean by more than a specified amount.

To make the problem more specific, suppose we wish to estimate the mean μ of a normal population and require $100(1 - \alpha)\%$ confidence that the estimate will be within a margin of d units of μ . In other words, for specified α and d , we want to find a sample size n and construct a $100(1 - \alpha)\%$ two-sided confidence interval in the form of $\bar{Y} \pm d$. There are two cases to consider: σ known and σ unknown.

When σ is known (an unlikely case), then, from Equation (9.10), the required confidence interval is $\bar{Y} \pm Z_{\alpha/2}\sigma / \sqrt{n}$. We set the margin specification d equal to $Z_{\alpha/2}\sigma / \sqrt{n}$ and solve for n to find:

$$n = \left(\frac{z_{\alpha/2}\sigma}{d} \right)^2 = z_{\alpha/2}^2 \left(\frac{\sigma}{d} \right)^2 \quad (9.21)$$

If $\alpha = 0.05$, $z_{\alpha/2} = 1.96$ and Equation (9.21) becomes:

$$n = 1.96^2 \left(\frac{\sigma}{d} \right)^2, \text{ rounded to } 4 \left(\frac{\sigma}{d} \right)^2 \quad (9.22)$$

Since σ is known and d is given, we substitute those values in Equation (9.22) to obtain the required sample size.

Example 9.5. Travel voucher error. We wish to calculate the number of travel vouchers to sample to estimate average voucher error (either over- or under-reported) within \$10, all with 95% confidence. Assume that the voucher error is distributed normally with a known $\sigma = \$50$.

Using Equation (9.22), we have:

$$n = 4 \left(\frac{50}{10} \right)^2 = 100$$

Even if σ is not known, we can still use Equation (9.22) if the required margin of error is specified as a multiple of σ . In other words, if $r = d/\sigma$ is given, we can calculate the required sample size. For instance, the same sample size of $n = 100$ in Example 9.5 would have been obtained if all we are given is that $r = 0.2$.

9.15 Sample size determination: σ unknown

Even if we do not know the standard deviation σ , we can still achieve the error margin d with $100(1 - \alpha)\%$ confidence as in Section 9.14 by using a two-stage approach. In this approach, we

Stein's procedure first estimate σ with a preliminary sample and then draw a second sample so that the required error margin can be achieved with the combined sample. This approach is called **Stein's procedure** (or *Stein's two-stage procedure*), as discussed and slightly modified by Desu and Raghavarao (1990), p. 4.

Stage 1: Take an initial random sample of size n_1 of at least three observations. Calculate the sample mean, denoted by \bar{Y}_1 , and the sample standard deviation, denoted by S_1 .

- Although we have considerable liberty in selecting the initial sample size, we should avoid taking an initial sample that is too small. The idea is to choose n_1 so as to get a reasonable estimate of σ without having n_1 larger than necessary to meet our requirements.

Next, use Table T-3 of the appendix to find $t_{(1-\alpha/2)}(n_1 - 1)$. Determine n_0 from Equation (9.23):

$$n_0 = \left[S_1 t_{(1-\alpha/2)}(v_1) / d \right]^2 \quad (9.23)$$

where $v_1 = n_1 - 1$ degrees of freedom.

Stage 2: Let $n_2 = n_0 - n_1$. That is, we need to collect n_2 observations above and beyond the n_1 observations we have already collected in the first stage.

If $n_2 \leq 0$, no additional observations are required and \bar{Y}_1 is within the prescribed margin of error d (with the prescribed $100(1 - \alpha)\%$ confidence. Otherwise, make n_2 additional observations.

Denote the sample mean of the n_2 observations by \bar{Y}_2 .

Let \bar{Y} be the sample mean of the $n_1 + n_2$ observations. \bar{Y} may be calculated as:

$$\bar{Y} = \frac{n_1 \bar{Y}_1 + n_2 \bar{Y}_2}{n_1 + n_2} \quad (9.24)$$

Then \bar{Y} is an estimator that is within a margin of error d for estimating μ with $100(1 - \alpha)\%$ confidence.

Example 9.5, continued. In this part of the example, σ is not known, but we still want $d = \$10$. We follow the prescribed procedure:

Stage 1: We take an initial random sample of size $n_1 = 15$ observations. Suppose the sample yields $\bar{y}_1 = 30.43$ and $S_1 = 28.33$.

From Table T-3, we find $t_{0.975}(14) = 2.14$. So, $n_0 = [(28.33) (2.14) / 10]^2 = 36.76$, rounded up to 37.

Stage 2: We calculate $n_2 = n_0 - n_1 = 37 - 15 = 22$. Hence, we take 22 additional observations. Suppose the mean of that sample is $\bar{y}_2 = 38.07$. Now we calculate the mean of the combined sample of 37 observations:

$$\bar{y} = \frac{(15)(30.43) + (22)(38.07)}{15 + 22} = 34.97$$

Thus, we are 95% confident that our estimate $\bar{y} = 34.37$ does not differ by more than \$10 from the population mean.

See Desu and Raghavarao (1990) for an extended treatment of sample size determination.

10

Inference

10.1 What to look for in Chapter 10

hypothesis testing Statistical inference is based on a procedure called *hypothesis testing*. Chapter 10 defines hypothesis testing and explains the rationale behind it, thus building a foundation for specific material developed in later chapters. Among the terms employed in hypothesis testing are the following:

- *statistical significance, §10.2*
- *null hypothesis, H_0 , §10.3*
- *alternative hypothesis, H_1 , §10.3*
- *test statistic, §10.3*
- *critical point, §10.3*
- *critical and noncritical regions, §10.3*
- *power of a test, §10.3*

We will also encounter two types of errors that can occur in hypothesis testing:

- *Type I error, probability of Type I error, §10.3*
- *Type II error, probability of Type II error, §10.3*

10.2 Testing statistical hypotheses: setting the stage

statistically significant
significant

Statistical investigations look for statistical significance that may be demonstrated by the data. A finding from an investigation is called *statistically significant* or simply *significant* if it unlikely to have occurred by chance alone. For example, suppose you flip a coin 10 times, only to realize 10 disappointing tails. The sequence of 10 consecutive tails could have occurred as described by chance alone. However, if the coin is fair, the probability of a sequence of 10 tails is $(0.5)^{10}$, or about 0.001. This small probability could lead us to state that the sequence of tails we observed is significant. In this context, any event that is unlikely to occur is significant if the probability of its occurrence is small, relative to a preset determination of what constitutes small.

Statistical investigations (including experiments, procedures, tests, etc.) must be conducted in an orderly fashion. They must be designed carefully, executed as planned, analyzed correctly and interpreted properly. The steps in the conduct of the investigation must be understood and, preferably, written down and agreed upon by all stakeholders before the start of the actual investigation. The protocol should list the model for the data and all the assumptions made about the data structure. For a focused investigation, the objectives of the investigation should be written as a pair of hypotheses: a null hypothesis and an alternative hypothesis. These hypotheses are defined in Section 10.3.

Two seemingly unrelated examples explain the logic of testing statistical hypotheses. Example 10.1 involves a traffic violation. Example 10.2 examines response time to security violations.

Example 10.1. Traffic violation. We are driving our car leisurely when a police officer pulls us over and suggests that we just ran a stop sign. Adamantly, but respectfully, we disagree, contending that we did not do such a thing. Just as adamantly, the officer holds an alternative opinion and invites us to traffic court to resolve the matter. *How do we deal with this problem? Indeed, what is the problem? Is there a hypothesis to be tested? What does it have to do with statistics anyhow?*

Example 10.2. Response time to intrusions. Suppose agency regulations require that a security system should pinpoint the location of peripheral intrusions within an average of 10 seconds. To investigate whether the system is in compliance, the system is tested by a number of random provocations. At each such provocation, the system's response time is recorded, and the average response time to these provocations is calculated. The observed average response time is compared with the regulatory requirement, and a statistical analysis is performed to determine if the system is in compliance.

How do we deal with this problem? Indeed, what is the problem? Is there a hypothesis to be tested? What do Examples 10.1 and 10.2 have in common?

Consider next the following parallel statements comparing traffic court processes with statistical hypothesis testing:

Table 10.1. Traffic court processes and hypothesis testing

Example 10.1	Example 10.2
<p>Traffic courts in the United States work under a process where evidence is examined to see whether it contradicts the accused’s claim of innocence. <i>What evidence is there to contradict the accused’s claim that he did not run a stop sign?</i></p>	<p>Statistical hypothesis testing is a process by which a set of data is examined to see whether the data contradict some statement about the population from which the data were drawn. <i>What data are there to contradict the claim that the average response time does not exceed 10 seconds?</i></p>
<p>The claim of innocence is a hypothesis about the accused. <i>One is innocent until proven guilty with a high degree of certainty (i.e., “beyond a reasonable doubt”).</i></p>	<p>The statement is a hypothesis about the population. <i>The claim that the system average response time is 10 seconds is accepted unless the data suggest otherwise.</i></p>
<p>The protocol for examining evidence in traffic courts is well defined; if applied properly, it is an orderly sequence of rational and defensible steps leading to a finding of guilty or not guilty. <i>We are assured that we can face the accuser, question evidence, examine the witnesses, and present evidence in our defense during the trial.</i></p>	<p>The protocol for statistical hypothesis testing is well defined; if applied properly, it is an orderly sequence of rational and defensible steps leading to rejection or nonrejection of the null hypothesis. <i>Data must be selected according to established rules; they must be analyzed according to a prescribed protocol, and the results must be properly interpreted.</i></p>

This parallel structure illuminates the logic of hypothesis testing. We see that there are strong similarities between court procedures, in which material is examined as evidence of a violation, and hypothesis testing, in which data are examined as evidence of the falseness of a claim about a population.

10.3 Terminology

- statistical hypothesis** A **statistical hypothesis** is a statement about a population. In most statistical investigations, this is a quantitative statement that can be written in mathematical notation.
- null hypothesis** A **null hypothesis** (symbolized by H_0) is a statement about a population's parameters, a function of those parameters, or the structure of the population. In Example 10.2 (response time), the null hypothesis about the population mean is understood to be $\mu \leq 10$ but is conventionally written as $H_0: \mu = 10$. (This book follows the usual practice of writing the null hypothesis as an equality.)
- The subscript 0 in H_0 (read and pronounced “zero” or “naught”) sometimes implies “no difference,” “no change,” or “status quo.”
- alternative hypothesis** An **alternative hypothesis** (symbolized by H_1) is a statement that contradicts a null hypothesis. In testing hypotheses, we accept H_1 if there is sufficient evidence to reject H_0 (i.e., if the data are inconsistent with H_0). An alternative hypothesis is usually stated as an interval. Thus, as in Example 10.2, if the null hypothesis is $H_0: \mu = 10$, then the alternative hypothesis could potentially be any one of the following: $H_1: \mu > 10$, $H_1: \mu < 10$, or $H_1: \mu \neq 10$. However, because the actual claim is that $\mu \leq 10$, the correct alternative hypothesis for Example 10.2 is $H_1: \mu > 10$.
- Some writers denote the alternative hypothesis by H_A .
- right-sided and left-sided hypotheses** An alternative hypothesis written with a greater than sign (e.g., $\mu > 10$) is a **right-sided hypothesis**. An alternative hypothesis written with a less than sign (e.g., $\mu < 10$) is a **left-sided hypothesis**. Both left-sided and right-sided alternative hypotheses are **one-sided alternative hypotheses**. An alternative hypothesis written with the not-equal symbol (e.g., $\mu \neq 10$) is a **two-sided hypothesis**.
- working hypothesis** In some disciplines, an alternative hypothesis is called a **working hypothesis** or a **research hypothesis** because this is the hypothesis the researcher usually hopes to validate. For example, if a new drug is investigated, the usual null hypothesis is that it is no different than the old drug (or a placebo). If the data are inconsistent with H_0 , the alternative hypothesis is accepted, and the happy researcher may be ready to write and publish his findings.
- research hypothesis**

test statistic A **test statistic** is a function of the data that is used to test a null hypothesis. A **hypothesis test** or **test of hypothesis** is a procedure that leads to rejection or non-rejection of the null hypothesis. In Example 10.2, the test statistic would be the standard normal variable:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma / \sqrt{n}} \quad (10.1)$$

where μ_0 is the value of μ specified by H_0 . The statistical test is detailed in Chapter 13.

In a courtroom situation, the element analogous to a test statistic is a witness, such as a bystander or police officer, or a recording device.

level of significance A **level of significance** (symbolized by α) is a measure of the risk, quantified as a probability, which we are willing to take in rejecting H_0 when in actuality H_0 is correct. In most work at the NRC, α is set at 0.05. In this book, unless otherwise specified, we use $\alpha = 0.05$ as a default value.

- The choice of α is somewhat arbitrary. Studies and investigations that set α at 0.05 often do so because it is “conventional,” and there is no compelling reason to choose another level of significance. Historical (and, perhaps, psychological) reasons for choosing this level of significance can be found in the literature. One such source is a paper by Stiegler (2008), which traces the selection of this level to R. A. Fisher, the most prominent founder of modern statistics.

critical region Given a test statistic, a **critical region** (also called a **rejection region**) is the set of all possible values of the test statistic that lead to the rejection of H_0 , and, hence, the acceptance of H_1 . The critical region should be chosen to be consistent with the alternative hypothesis. For one-sided alternatives, the best critical region is a one-sided interval. Thus, for $H_1: \mu > \mu_0$, the critical region is $\bar{y} > c_1$ for some constant c_1 , and for $H_1: \mu < \mu_0$, the critical region is $\bar{y} < c_2$. For two-sided alternatives, the best critical region consists of two one-sided intervals. Thus, for $H_1: \mu \neq \mu_0$, the critical region is $\{\bar{y} < c_3, \bar{y} > c_4\}$.

noncritical region A **noncritical region** is the complement of a critical region. It is the set of all possible values of the test statistic that do not lead to the rejection of H_0 .

critical point or a **critical value** is any point that separates the critical region from the noncritical region. A test of hypothesis may involve more than one critical point. Thus, neither a critical region nor a noncritical region is necessarily contiguous. For example, if the alternative hypothesis is two-sided, there are two critical points, c_3 and c_4 , because we reject H_0 if \bar{y} is smaller than c_3 or larger than c_4 .

All of the terminology introduced so far deals only with performing hypothesis testing. We now introduce the terminology dealing with its evaluation.

Because the null hypothesis can be either true or false, we will commit an error if we draw the wrong conclusion from our hypothesis testing. We distinguish two types of errors, depending on the true underlying state of affairs. These errors are shown in Table 10.2. In some disciplines, this is sometimes called a **truth table**.

Table 10.2. Possible outcomes of a test of hypothesis

Statistical decision	Reality	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Type I error, false alarm, false positive A **Type I error** is the error committed if we reject the null hypothesis when, in fact, it is true. In some disciplines, a Type I error is also called a **false alarm** or a **false positive**.

In Example 10.1, if we are judged guilty when we did not actually run a stop sign, the judge committed a Type I error.

In Example 10.2, if the true average response time is no higher than 10 seconds but the sample happened to reject the null hypothesis, we committed a Type I error.

probability of a Type I error The **probability of a Type I error** is denoted by α . This is the same α as the level of significance. This probability is usually set at our (or our manager's) discretion. As stated above, the default value used by the NRC for α is 0.05 (5%).

probability of exceedance If the alternative hypothesis is right-sided, α is sometimes called the **probability of exceedance**, because it is the probability that the test statistic will exceed the critical value. The term “probability of exceedance” is sometimes extended to denote the probability of breaching any critical point, even if a left-sided or a two-sided alternative hypothesis is involved.

Type II error A **Type II error** is the error committed if we do *not* reject the null hypothesis when, in fact, it is false (i.e., when the alternative hypothesis is true). In some disciplines, a Type II error is called a **false negative**.

In Example 10.1, if we are not convicted when we actually did run a stop sign, the judge committed a Type II error.

In Example 10.2, if the true average response time is higher than 10 seconds but the sample did not reject the null hypothesis, we committed a Type II error.

probability of a Type II error The **probability of a Type II error** is denoted by β . Because the alternative hypothesis allows for many different values of the parameter (call it θ) being tested, β will be a function of θ . This is in contrast to the probability α of a Type I error, that depends only on the single value for θ specified by the null hypothesis. As a function of θ , β is called the **operating characteristic (OC)** of the test.

consumer's risk In business, β is sometimes called the **consumer's risk**.

power of the test The probability $1-\beta$ of rejecting the null hypothesis as a function of θ is called the **power of the test**. In addition to accepting the null hypothesis when it is true, we want our hypothesis test to reject the null hypothesis when it is false. In other words, we want the test to be sensitive (“powerful”) to the falseness of the null hypothesis.

Chapter 13 illustrates the calculation of the power and the operating characteristic of the test.

Surely you have noticed some imbalance in the treatment of the null and the alternative hypotheses: We accept H_1 when the data suggest that H_0 be rejected, but we do not accept H_0 when the data do not suggest rejection. Here is a brief explanation. If H_0 is not rejected (i.e., if there is insufficient evidence to reject H_0), that does not mean that H_0 is proven correct. It means only that the statistical evidence was not strong enough to reject H_0 . In other words, we can never prove the null hypothesis; we can only disprove it.

We don't always practice what we preach. We sometimes say that we "accept the null hypothesis" when we fail to reject it, although statisticians try to avoid this expression. Even though we know that to prove the truth of H_0 would require infinite data, we often act as though H_0 is true simply because we did not reject it. However, when we do this, we must be convinced that the sample was large enough so that H_0 had a reasonable chance of being rejected if it were false.

10.4 Null and alternative hypotheses: examples

This section gives examples of competing hypotheses (null and alternative) in various environments and points to the chapters where the statistical tests are developed.

Testing a single mean (Chapter 13)

Claim: Mean response time to an alarm is no more than 10 seconds.

$$H_0: \mu = 10 \quad H_1: \mu > 10$$

Claim: Average percent of fuel enrichment is at least 78%.

$$H_0: \mu = 78 \quad H_1: \mu < 78$$

Claim: Average piston displacement is 2 millimeters.

$$H_0: \mu = 2 \quad H_1: \mu \neq 2$$

Claim: Average radiation exposure during time t is no more than 0.10 rads.

$$H_0: \mu = 0.10 \quad H_1: \mu > 0.10$$

Claim: Control rods are activated within 3 seconds of demand.

$$H_0: \mu = 3.0 \quad H_1: \mu > 3.0$$

Comparing two means (Chapter 15)

Claim: The average readings of two instruments are the same.

$$H_0: \mu_A = \mu_B \quad H_1: \mu_A \neq \mu_B$$

Claim: The average salinity of tank A is no higher than that of tank B.

$$H_0: \mu_A = \mu_B \quad H_1: \mu_1 > \mu_B$$

Comparing several means (Chapter 16)

Claim: All four laboratories yield the same average reading.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad H_1: \text{not all } \mu\text{'s are equal}$$

or $H_1: \mu_i \neq \mu_j, \text{ some } i, j$

Claim: Average construction time for nuclear plants is the same for countries A, B, and C.

$$H_0: \mu_A = \mu_B = \mu_C \quad H_1: \text{not all } \mu\text{'s are equal}$$

Testing slope (Chapter 18)

Claim: The coefficient of expansion of this metal is the same as that of gold (14.43×10^{-6} per $^{\circ}\text{C}$).

$$H_0: \beta = 14.43 \times 10^{-6} \quad H_1: \beta \neq 14.43 \times 10^{-6}$$

Claim: The net increase in net count produced by nondestructive assay measurement per gram of material is equal to 70.1.

$$H_0: \beta = 70.1 \quad H_1: \beta \neq 70.1$$

Testing correlation (Chapter 19)

Claim: The cost of fuel at the pump and the price of tea in China are not correlated.

$$H_0: \rho = 0 \quad H_1: \rho \neq 0$$

Contingency table analysis (Chapter 12)

Claim: Gender (G) and promotion (P) are independent.

$$H_0: Pr\{G\&P\} = (Pr\{G\})(Pr\{P\})$$

$$H_1: Pr\{G\&P\} \neq (Pr\{G\})(Pr\{P\})$$

10.5 Consequences of hypothesis testing

Because the amount of data is always finite, the possibility of making an error in hypothesis testing is always present. Table 10.3 summarizes the consequences of statistical hypothesis testing in light of this possibility. The first column in

that table lists the consequences of rejecting the null hypothesis, while the second column lists the consequences of *not* rejecting it.

Table 10.3. Consequences of hypothesis testing

After the hypothesis is tested and...	
...if the test statistic falls into the critical region, we reject the null hypothesis H_0if the test statistic does <i>not</i> fall into the critical region, we do not reject the null hypothesis H_0 .
The story told by the data is not consistent with H_0 .	The story told by the data is consistent with H_0 .
We may have made an error in rejecting H_0 .	We may have made an error in not rejecting H_0 .
If H_0 is true, we committed a Type I error.	If H_0 is false, we committed a Type II error.
The probability of committing a Type I error, α , is “small”; we know its value (typically 5%), and we are prepared to live with this small chance.	The probability of committing a Type II error, β , is not known but we hope it’s small. β is a function of the departure from the null hypothesis (i.e., the specific value of the parameter under the alternative hypothesis).

We cannot have both α and β be small, as there is a balance between the two. If we allow α to increase, we automatically decrease the size of β , and vice versa. The question then becomes how much of a larger probability of a Type I error are we willing to tolerate to get a smaller β . Another option is to increase the sample size while keeping α constant because our statistical test will then become more sensitive to departures from the null hypothesis. But, regardless of the sample size, the proper balance between α and β as a function of the separation between the null and alternative hypotheses will always be an issue.

10.6 Guilty until found innocent

A test of hypothesis should be sensitive to an incorrect null hypothesis. In statistical terms, the test should have a reasonably small β , even if we don’t know its actual size. The sensitivity (power) of the test is influenced by several factors, one of which is the sample size, n . For a fixed α and a fixed alternative to H_0 , the probability of rejecting H_0 is an increasing function of n . In other words, the power increases as n increases. If, in Example 10.2, the Director of

Security does not want the null hypothesis ($H_0: \mu = 10$) to be rejected, it is to his advantage to have n be as *small* as possible.

- Note also that it would serve the Director's purpose to have the variability in the system as *large* as possible so that any departures from the null hypothesis would be lost in the noise.

In the interest of ensuring a quick response system, we may reverse the roles of H_0 and H_1 . Now we claim that $\mu = 10.1$ (say) and ask the Director of Security to reject this hypothesis in favor of $H_1: \mu < 10.1$. Now it is to the Director's advantage to have a large sample size and a low variability.

In a similar vein, consider the following example: A processing plant for nuclear material measures and records its inventory and compares it to the book records. Discrepancies between the two can be attributed to random fluctuations resulting from sampling errors, loss due to processing, or even loss or diversion. If we set $H_0 = 0$ to reflect random fluctuations and no loss, real losses may be undetectable with potentially dire consequences.

A switch between the traditional H_0 and H_1 addresses this problem. We select a quantity of concern Q (say, 1 kilogram) and require the plant to demonstrate that if a quantity Q of material is unaccounted for, corrective action would be taken. Statistically, we set $H_0 = Q$ and $H_1 < Q$.

Another example is the release of a decommissioned area after scrubbing. We claim (H_0) that the area's radiation level is at some permitted level, and the licensee must demonstrate that the contamination level is significantly below that level.

10.7 Finally...

The framework for testing statistical hypotheses involves designing the experiment, selecting the sample size, drawing the sample, specifying the hypotheses, selecting a test statistic, running the analysis, reporting the results, and interpreting the findings. If any of these elements is inappropriate or inadequate, the findings may be faulty.

We also assume that the data are collected, recorded, and transmitted correctly. We assume that no data laundering takes place and that care is taken to protect the integrity of the data at the source, at the recording station, during transmission, and during the analysis.

Hypothesis testing is very closely related to the construction of confidence intervals covered in Chapter 9. Consider, for example, a two-sided confidence interval (L, U) for μ . This interval is a set of plausible values of μ consistent with the data. Any value outside (L, U) is considered an unlikely value of μ . It

turns out that a test of hypothesis at the same level of significance will reject the null hypothesis if and only if the sample mean does not fall in the confidence interval (L, U) . The same result holds for a one-sided confidence interval. If the sample mean does not fall in the confidence interval, the corresponding null hypothesis will be rejected and vice versa. In this sense, hypothesis testing and confidence interval construction are two sides of the same coin; for each test of hypothesis, there is an equivalent confidence interval and vice versa.

11

Goodness-of-fit tests

11.1 What to look for in Chapter 11

goodness-of-fit test A *goodness-of-fit test* is a procedure that investigates whether a sample could have come from a specified distribution. Chapter 11 presents several goodness-of-fit tests and discusses their applicability. Terminology used in this presentation includes:

- *empirical distribution function, §11.7*

Four goodness-of-fit tests are examined in detail:

- *chi-square test for discrete distributions, §11.3*
- *chi-square test for normality, §11.5 and §11.6*
- *Kolmogorov-Smirnov test for continuous distributions, §11.8*
- *Shapiro-Wilk test for normality for small samples, §11.9*
- *D'Agostino test of normality for moderate and large samples, §11.10*

11.2 Testing goodness-of-fit

test of fit A goodness-of-fit test, often referred to as a *test of fit*, is designed to examine whether a sample has come from a postulated distribution. Here are a few examples where assumptions about the structure of the distribution are made and where a test of fit may be applicable.

- In rolling a die, we wish to test whether we play with a fair die, that is, whether all faces of the die are equally likely to come up. Similarly, in playing a casino's roulette wheel, we test whether, when played for a long time, each of the 38 possible outcomes comes up with essentially the same frequency.
- In grading a class on the "bell curve," we wonder whether the proportions of A's, B's, C's, D's, and F's are consistent with the proportions expected from the normal distribution.
- In selecting employees for drug testing, we wish to test whether the random number generator that is used for employee selection follows the discrete uniform distribution.
- In testing hypotheses about the population mean, we sometimes make an assumption that the sample came from a normal distribution and ask whether this is a reasonable assumption.

The following notations are used:

- $F^*(y)$ is the hypothesized, or postulated, cumulative distribution function.
- $F(y)$ is the actual cumulative distribution function that gave rise to the sample.
- $F_n(y)$ is the cumulative distribution function that is constructed from the sample of size n (section 11.7).
- $f^*(y)$ is the hypothesized, or postulated, probability function.
- $f(y)$ is the actual probability function of the distribution that gave rise to the sample.
- $f_n(y)$ is the probability function that is constructed from the sample of size n .

Suppose we have a sample from $F(y)$ —or, equivalently, from $f(y)$ —and we wish to test whether $F(y)$ is the same as $F^*(y)$ —or, equivalently, if $f(y)$ is the same as $f^*(y)$. Formally stated:

$$\begin{aligned} H_0: F(y) = F^*(y) \quad \text{or} \quad f(y) = f^*(y) \\ H_1: F(y) \neq F^*(y) \quad \text{or} \quad f(y) \neq f^*(y) \end{aligned} \tag{11.1}$$

Intermediate steps in testing H_0 may involve one of the following:

- Construction of a histogram $f_n(y)$ from the data to mimic $f(y)$ (similar to that which is shown in Figure 6.2).
- Construction of a cumulative distribution function histogram $F_n(y)$ from the data to mimic $F(y)$ (similar to that shown in Figure 6.3).

empirical distribution function $F_n(y)$ is called the **empirical distribution function**. Section 11.7 outlines the construction of the empirical distribution function.

In this chapter, we present several tests of fit. Some of the tests are quite general in that they can apply to just about any distribution, while other tests are more specific, such as tests that apply only to the normal distribution. Also, some tests apply to distributions that are completely specified (i.e., with all parameters given), whereas other tests are applicable even when not all parameters are known and specified.

Among the methods for testing goodness-of-fit, some are superior to others in their sensitivity to different types of departures from the hypothesized distribution. For example, some tests are more likely to detect discrepancies between $F(y)$ and $F^*(y)$ when the discrepancy is present in the neighborhood of the center of $F(y)$, while other tests are more sensitive to a discrepancy around the tail of the distribution. Thus, we may not always choose the best test since we don't always know where the two distributions are likely to differ.

In the following sections, we present several of the most widely used tests of fit. However, it must be emphasized that, in order for the statistical statements about the tests to hold, only one test should be used, and the selection of that test should be made before the data are seen.

Finally, like every test of hypothesis, a test of fit can at best indicate, at a prespecified significance level, that if the null hypothesis is rejected, the sample is not from the hypothesized distribution. However, if the null hypothesis is not rejected, this does not mean that the sample necessarily came from the hypothesized distribution. It usually means that it is reasonable to proceed as though the hypothesis were true.

11.3 Chi-square test for discrete distributions

**chi-square
goodness-of-
fit test**

The **chi-square goodness-of-fit test** compares frequencies of sample values with the frequencies of a hypothesized probability function $f^*(y)$. This test usually involves grouping the sample data into predetermined and mutually exclusive classes (sometimes called bins) which cover the full range of the population from which the sample is drawn. For each class, the number of observed data values is compared to the number that is theoretically expected from $f^*(y)$. Practical considerations determine the number of classes in a test. Such considerations are discussed in the context of the examples we present.

- The chi-square test of fit is a popular test because of its relative ease of calculation. On the other hand, the chi-square test has several shortcomings. First, the grouping of the data is not unique, and different groupings can lead to different conclusions about the agreement between the data and $f^*(y)$. Second, the grouping of the data may lead to a loss of detail and hence to some loss of sensitivity to departures from $f^*(y)$.

**chi-square
statistic**

Given a sample of size n , we wish to test the null hypothesis $H_0: f(y) = f^*(y)$ against a two-sided alternative, where $f(y)$ is the probability function of the distribution that gave rise to the sample and $f^*(y)$ is a hypothesized probability function. We group the sample data into c classes and use a test statistic called the **chi-square statistic**, denoted by $\chi^2(c-1)$, calculated as:

$$\chi^2(c-1) = \sum_{i=1}^c \frac{O_i - E_i}{E_i}^2 \quad (11.2)$$

where:

i = class index

O_i = count of observations in class i

E_i = expected count in class i

c = number of classes

**expected
count**

The **expected count** E_i in class i is n times the probability of an observation falling in class i , based on the hypothesized probability function $f^*(y)$. Note that E_i is not necessarily an integer.

A mathematically equivalent form of Equation (11.2) is given by:

$$\chi^2(c-1) = \sum_i^c \frac{O_i^2}{E_i} - n \quad (11.3)$$

Although Equations (11.2) and (11.3) are algebraically equivalent, Equation (11.3) is usually easier to calculate. However, the form of Equation (11.2) is more useful because it emphasizes that the chi-square test is based on the differences between the two distributions.

Under the null hypothesis, $\chi^2(c-1)$ has a chi-square distribution with $c-1$ degrees of freedom (see Section 7.10).

Example 11.1. Rolling a die. In an experiment designed to test whether a specific die is fair, the die is rolled $n = 600$ times. If the die is fair, then the probability π_i of showing face i ($i = 1, 2, \dots, 6$) is $1/6$ and the expected count (also called the *expected frequency*) in class i is $(n)(\pi_i) = (600)(1/6) = 100$. The null and alternative hypotheses and the level of significance are:

$$H_0: f(i) = \pi_i = 1/6, i = 1, 2, \dots, 6$$

$$H_1: f(i) = \pi_i \neq 1/6, \text{ for some } i$$

$$\alpha = 0.05$$

- The alternative hypothesis means that not all π_i are equal (i.e., at least two π_i 's, are different).

In each of the 600 rolls, the number of the face showing is recorded. Table 11.1 gives the class counts.

Table 11.1. Observed and expected frequencies of rolling a die

Face (dots), i	1	2	3	4	5	6	Total
Observed count, O_i	123	131	84	95	96	71	600
Expected count, E_i	100	100	100	100	100	100	600
Difference, $O_i - E_i$	23	31	-16	-5	-4	-29	0

The calculated chi-square statistic, using Equation (11.2) with $c-1 = 5$ degrees of freedom, is:

$$\chi^2(5) = \frac{23^2}{100} + \frac{31^2}{100} + \frac{(-16)^2}{100} + \frac{(-5)^2}{100} + \frac{(-4)^2}{100} + \frac{(-29)^2}{100} = 26.28$$

From Table T-2 of the appendix, the critical point for $\alpha = 0.05$ is $\chi_{0.95}^2(5) = 11.10$. Since $\chi^2(5) > 11.07$, we reject the hypothesis of equal probability. We have statistical evidence that the die is not fair ($p < 0.05$). Note that the statement " $p < 0.05$ " indicates that the test is significant at the 0.05 level.

- The conclusion may also be worded to state that the sample proportions differ significantly from the hypothesized proportions.

Example 11.2. Grading a class. School policy suggests that overall grades throughout the campus be awarded in the following proportions: A: 10%, B: 20%, C: 55%, D: 10%, F: 5%. A class of 81 students was graded, and the administration wonders whether the grades for the class are consistent with school policy.

The null and alternative hypotheses are written as:

$$H_0: \pi_A = 0.10; \pi_B = 0.20; \pi_C = 0.55; \pi_D = 0.10; \pi_F = 0.05$$

H_1 : At least one of the stated proportions is not met.

Table 11.2 gives the data obtained from the instructor.

Table 11.2. Observed and expected count in class grades

Grade	A	B	C	D	F	Total
Observed count, O_i	4	8	54	10	5	81
Expected proportion, π_i	0.10	0.20	0.55	0.10	0.05	1.00
Expected count, E_i	8.10	16.20	44.55	8.10	4.05	81.00
Difference, $O_i - E_i$	-4.10	-8.20	9.45	1.90	0.95	.00

The calculated chi-square statistic, per Equation (11.2), is:

$$\chi^2(4) = \frac{(-4.10)^2}{81} + \frac{(-8.20)^2}{81} + \frac{9.45^2}{81} + \frac{1.90^2}{81} + \frac{0.95^2}{81} = 2.20$$

From Table T-2 of the appendix, the critical point for $\alpha = 0.05$ is $\chi_{0.95}^2(4) = 9.49$. Since $\chi^2(4) = 2.20 < 9.49$, we have no reason to claim that school policy was violated. We conclude that random fluctuations most likely explain the grade distribution for this class.

Example 11.3. Random selection. Random selection of employees for drug testing is made using a computer routine, which randomly selects $n = 150$ out of $N = 1,491$ employees for the test. Responding to numerous employee complaints, the selection system is tested for randomness. For this test, we arbitrarily divide the employees into seven groups according to their last name. The null and alternative hypotheses are similar to those of Example 11.2. Table 11.3 shows the count of the employees in each group in the agency and in the sample and the relevant calculations.

Table 11.3. Testing randomness in employee selection

Last name	A-C	D-G	H-K	L-N	O-R	S-W	X-Z	Total
Agency count	310	253	177	163	244	268	76	1,491
Agency proportion, π_i	0.21	0.17	0.12	0.11	0.16	0.18	0.05	1.00
Observed count, O_i	28	24	19	23	30	16	10	150
Expected count, E_i	31.19	25.45	17.81	16.40	24.55	26.96	7.65	150
Difference $O_i - E_i$	-3.19	-1.45	1.19	6.60	5.45	-10.96	2.35	0.00
$(O_i - E_i)^2$	10.16	2.11	1.42	43.58	29.73	120.16	5.54	
$(O_i - E_i)^2 / E_i$	0.33	0.08	0.08	2.66	1.21	4.46	0.72	9.54

From Table 11.3, the chi-square statistic is calculated as $\chi^2(6) = 9.54$. From Table T-2 of the appendix, the critical point for $\alpha = 0.05$ is $\chi_{0.95}^2(6) = 12.6$. Since the critical point is not exceeded, we have no reason to claim that the selection is not made at random.

11.4 Chi-square test: sample size considerations

The use of the chi-square statistic to test goodness-of-fit may be inappropriate for small sample sizes. Recommended rules for tackling this problem exist in the statistical literature. They often include cautions about interpreting the results. These rules are stated in terms of the classes' expected counts rather than the number of observed counts. One such widely used rule, known as the "Rule of 5," states that no class should have a smaller expected count than 5. Because there must be at least two classes, this rule implies that the sample size must be at least 10. To satisfy the rule, the data must be partitioned so that the sample size multiplied by the smallest class proportion is no smaller than 5.

Some authors offer alternatives to the Rule of 5. For example, Dixon and Massey (1983), p. 277, allow chi-square tests that meet either one of two conditions:

- Condition 1: No expected count is smaller than 2 for any class.
- Condition 2: At least 80% of all the classes have expected counts of at least 5, with none of the expected counts less than 1.

There are two strategies to use if a rule is not satisfied. One is to increase the sample size, and the other is to consolidate classes. It is preferable to increase the sample size because this will increase the sensitivity of the test to detect departures from the null hypothesis. Consolidating classes will decrease that sensitivity.

11.5 Chi-square test for normality: known parameters

The chi-square test applies equally well to data from a continuous distribution. This section illustrates the use of the chi-square test for testing normality when the hypothesized normal distribution is completely specified (i.e., the parameters μ and σ^2 are given). Section 11.6 will consider the case where the parameters are not given and are estimated from the sample.

In both cases, the sample data are sorted into classes, and the chi-square statistic is calculated from the difference between the expected and observed counts. We illustrate the use of the chi-square test of fit by the following example, borrowed from Bowen and Bennett (1988), p. 525.

Example 11.4. Uranium content of air filters. A fuel fabrication facility routinely monitors its gaseous effluents by using an alpha counter to measure the uranium content of air filters. The alpha counter output is the number of counts observed in 1 minute, which can be translated into a quantity of uranium given the appropriate calibration relationship. The facility purchased a new alpha counter and must test the manufacturer's claim that the counter's random measurement errors follow a normal distribution with a standard deviation of 10 counts/minute for samples yielding between 200 and 400 counts/minute. A standard rated at 310 counts/minute was measured 30 times, yielding the counts/minute arranged in ascending order of magnitude, as shown in Table 11.4.

Table 11.4. Uranium count per minute

267	270	278	280	283	284	287	296	299	300
300	304	306	313	314	314	315	316	316	316
317	317	319	321	325	326	331	342	351	354

Given $\mu = 310$ and $\sigma^2 = 100$, we proceed to test:

$$H_0 : Y \sim N(310, 100)$$

$$H_1 : Y \not\sim N(310, 100)$$

where the symbol \sim is read as “distributed as” and the symbol $\not\sim$ indicates “not distributed as.”

We elect to divide the domain of the hypothesized distribution into four equally likely intervals, where the split points for the intervals are the quartiles. Interpolating in Table T-1 of the appendix, we find $z_{0.25} = -0.674$, $z_{0.50} = 0.000$, and $z_{0.75} = 0.674$. Thus, the quartiles of the hypothetical distribution are:

$$y_{.25} = 310 + (z_{.025})(10) = 310 + (-0.674)(10) = 303.3$$

$$y_{.50} = 310 + (z_{.50})(10) = 310 + (0.00)(10) = 310.0$$

$$y_{.75} = 310 + (z_{.75})(10) = 310 + (0.674)(10) = 316.7$$

Table 11.5 shows the calculation of the chi-square statistic with $c - 1 = 3$ degrees of freedom.

Table 11.5. Testing normality: known parameters

Class:	< 303.3	303.3 – 310.0	310.0 – 316.7	> 316.7	Total
Observed count, O_i	11	2	7	10	30
Expected count, E_i	7.5	7.5	7.5	7.5	30
Difference, $O_i - E_i$	3.5	-5.5	-0.5	2.5	0.0
$(O_i - E_i)^2$	12.25	30.25	0.25	6.25	
$(O_i - E_i)^2 / E_i$	1.63	4.03	0.03	0.83	6.53

From Table 11.5, the calculated chi-square is 6.53. From Table T-2 of the appendix, the critical point for $\alpha = 0.05$ is $\chi_{0.95}^2(3) = 7.81$. Because the critical

point is not exceeded, we have no statistical evidence to reject the hypothesized distribution at a 5% significance level.

- Had we chosen $\alpha = 0.10$, H_0 would have been rejected because $\chi_{0.90}^2(3) = 6.25$, which is exceeded by the calculated chi-square. However, because H_0 states that the data come from a normal distribution with specified mean and variance, it is possible that the data are from a normal distribution with a different mean and/or variance. This possibility is considered in Section 11.6.

11.6 Chi-square test for normality: unknown parameters

Parameters of a distribution are rarely known or specified. Hence the chi-square test for normality may not be applicable, at least not in the format presented, as both normal parameters are used in the determination of the class boundaries. We continue with the data in Example 11.4, this time without having the mean or the variance specified.

Example 11.5. Uranium content of air filters: unknown parameters.

The data are the same as in Table 11.4, except that the information did not come from measuring a standard with a known mean. We test the null hypothesis that the data came from a normal distribution with unspecified mean and variance.

Without knowledge of μ and σ , class boundaries cannot be tied to quantiles of the normal distribution as was done in Section 11.5. Instead, we begin by arbitrarily grouping the data in Table 11.4 into four classes as shown in Table 11.6, where the midpoint of each class is also shown. Note that these classes do not cover the full range of the normal distribution. The only requirement is that the classes cover the full range of the sample data in Table 11.4.

Table 11.6. Class selection

Class	Class boundaries	Class midpoint, m_i	Observed count, o_i
1	(260, 285)	272.5	6
2	(285, 310)	297.5	7
3	(310, 335)	322.5	14
4	(335, 360)	347.5	3

Next, as in Section 2.4, we calculate the sample weighted mean as:

$$\bar{Y}_W = \sum_{i=1}^c O_i m_i / n \quad (11.4)$$

$$= [6(272.5) + 7(297.5) + 14(322.5) + 3(347.5)] / 30 = 309.167$$

and the sample weighted variance as:

$$S_W^2 = \sum_{i=1}^c O_i m_i^2 / n - \bar{Y}_W^2 \quad (11.5)$$

$$= [6(272.5)^2 + 7(297.5)^2 + 14(322.5)^2 + 3(347.5)^2] / 30 - (309.167)^2$$

$$= 530.35$$

from which the standard deviation S_W is calculated as $s_W = 23.03$.

We now use the estimated mean and standard deviation instead of specified parameters to calculate the chi-square test statistic as in Section 11.5. First, we add a lower class $(-\infty, 260)$ and an upper class $(360, \infty)$ to the four classes in Table 11.6. The lower class is designated Class 0 and the upper class as Class 5, with the four classes in Table 11.6 becoming Class 1 – 4. Table 11.7 lists the class boundaries.

Table 11.7. Class boundaries and expected counts

Class	Class boundaries	Probability, π_i	Expected count, E_i
0	≤ 260	0.0164	0.49
1	(260, 285)	0.1306	3.92
2	(285, 310)	0.3674	11.02
3	(310, 335)	0.3546	10.64
4	(335, 360)	0.1173	3.52
5	> 360	0.0137	0.41

Next, the proportion π_i assigned to each class is calculated. For Class 1, we have:

$$\begin{aligned}\pi_1 &= Pr\{260 < Y \leq 285\} \\ &= Pr\left\{\frac{260 - 309.167}{23.03} < Z \leq \frac{285 - 309.167}{23.03}\right\} \\ &= Pr\{-2.135 < Z \leq -1.049\} \\ &= Pr\{Z < -1.049\} - Pr\{Z \leq -2.135\} \\ &= (1.0000 - 0.8530) - (1.0000 - 0.9836) \\ &= 0.1470 - 0.0164 = 0.1306\end{aligned}$$

from which the expected count in Class 1 is calculated as $(.1306)(30) = 3.92$.

Similarly, we find that:

$$\pi_2 = Pr\{285 < Y \leq 310\} = 0.3674. \text{ The expected count for Class 2 is } 30(0.3674) = 11.02.$$

$$\pi_3 = Pr\{310 < Y \leq 335\} = 0.3546. \text{ The expected count for Class 3 is } 30(0.3546) = 10.64.$$

$$\pi_4 = Pr\{355 < Y \leq 360\} = 0.01173. \text{ The expected count for Class 4 is } 30(0.1173) = 3.52.$$

Table 11.7 summarizes these calculations. Note that the sum of the expected counts for the four classes is 29.10. The missing expected count of 0.90 stems from the omission of the expected counts for Class 0 and Class 5. The probabilities of falling in these end classes are:

$$\begin{aligned}Pr\ Y \leq 260 &= Pr\left\{Z \leq \frac{260 - 309.167}{23.03}\right\} Pr\{Z \leq -2.1346\} = 0.0164 \\ Pr\ Y > 360 &= Pr\left\{Z > \frac{360 - 309.167}{23.03}\right\} Pr\{Z > 2.2069\} = 0.0137\end{aligned}$$

The expected counts for Class 0 and Class 5 are $30(.0164) = .49$ and $30(.0137) = .41$, respectively. The calculations for Table 11.7 are now complete. Because the expected counts for the end classes are so small, we combine Class 0 with Class 1 and Class 5 with Class 4. Table 11.8 shows the revised class boundaries, observed counts, and expected counts.

Table 11.8. Class boundaries and counts (revised)

Class	Class boundaries	Observed count, O_i	Expected count, E_i
1	$(-\infty, 285)$	6	4.41
2	$(285, 310)$	7	11.02
3	$(310, 335)$	14	10.64
4	$(335, \infty)$	3	3.93

The chi-square statistic is now calculated in Table 11.9.

Table 11.9. Testing normality: unknown parameters

Class	< 285	285–310	310–335	> 335	Total
Observed count, O_i	6	7	14	3	30
Expected count, E_i	4.41	11.02	10.64	3.93	30
Difference, $O_i - E_i$	1.59	-4.02	3.36	-.93	0.00
$(O_i - E_i)^2$	2.53	16.16	11.29	.86	
$(O_i - E_i)^2 / E_i$	0.57	1.47	1.06	0.22	3.32

The calculated chi-square statistic is 3.32.

When the parameters of the distribution are completely specified, the number of degrees of freedom for the chi-square test is $c - 1$, where c is the number of classes. When one or more parameters are not specified, the number of degrees of freedom is reduced by the number of parameters that are estimated. If k parameters are estimated from the sample, then the number of degrees of freedom is:

$$v = c - k - 1 \quad (11.6)$$

In this example, we have four classes and two estimated parameters. From Equation (11.6), we have $v = 1$. Hence, we look up the 0.95 quantile in Table T-2 of the appendix for $df = 1$ to get $\chi_{0.95}^2(1) = 3.84$. Since $3.32 < 3.84$, we do not have enough statistical evidence to reject normality at the 5% level of significance. However, had we set the level of significance at 10%, then the critical value would have been 2.71, and we would have rejected the null hypothesis of normality. One interpretation of these apparently contradictory results is to conclude that there is moderate but not strong statistical evidence that the sample did *not* come from a normal distribution. This is because

rejecting a null hypothesis at the 5% level of significance requires stronger statistical evidence than rejecting it at the 10% level.

11.7 Empirical cumulative distribution function

empirical cumulative distribution function Several goodness-of-fit tests for continuous distributions are based on an estimate of the underlying cumulative distribution function. Consider a sample $\{y_1, y_2, \dots, y_n\}$ of size n from a continuous distribution with *cdf* $F(y)$. The **empirical cumulative distribution function**, or **empirical cdf**, denoted by $F_n(y)$.

$$F_n(y) = [1/n][\text{number of observations that are less than or equal to } y] \quad (11.7)$$

To write Equation (11.7) in mathematical notation, we rearrange the n observations in the sample in ascending order and denote the i^{th} ordered observation by $y_{(i)}$ so that:

$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)} \quad (11.8)$$

- Note the subscripts in parentheses that are used to denote ordered observations. Some writers use a superscript, such as $y^{(i)}$, for the same purpose. Still other writers elect to use different letters to denote order statistics, thus avoiding parentheses around the subscripts. For example, they may use z_1 to denote the smallest observation, z_2 the second smallest, and so on.

rank The subscript i in $y_{(i)}$ assigned to the i^{th} ordered observation is called the **rank** of that observation.

Using the ordered observations, Equation (11.7) can be written as:

$$\begin{aligned} F_n(y) &= 0, \text{ if } y \leq y_{(1)} \\ &= k/n, \text{ if } y_{(k)} \leq y < y_{(k+1)}, k = 1, 2, \dots, n-1 \\ &= 1, \text{ if } y \geq y_{(n)} \end{aligned} \quad (11.9)$$

step function $F_n(y)$ is a **step function**, which is illustrated in Figure 11.1 for a set of 10 observations $\{8, 3, 7, 9, 9, 5, 8, 6, 11, 10\}$. We first order the observations in ascending order to yield the set $\{3, 5, 6, 7, 8, 8, 9, 9, 10, 11\}$. Then we proceed to plot $F_n(y)$ against y . Note that when r observations are tied, (like when $y_{(i)} = y_{(i)} = 8$), $F_n(y)$ steps up by r/n .

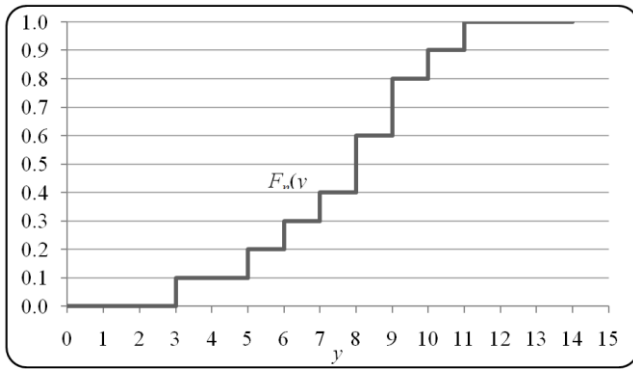


Figure 11.1 Empirical cumulative distribution function

The empirical *cdf* can be used to test whether a dataset came from a hypothesized distribution. For example, to test whether the 10 observations from Figure 11.1 came from the normal distribution with mean $\mu = 7.0$ and standard deviation $\sigma = 2.0$, we can superimpose the *cdf* for $N(7.0, 4.0)$ on Figure 11.1, as in Figure 11.2. To have a statistically valid goodness-of-fit test, it is necessary to quantify the graphical discrepancies between the two *cdfs* in Figure 11.2. The next section discusses one such test.

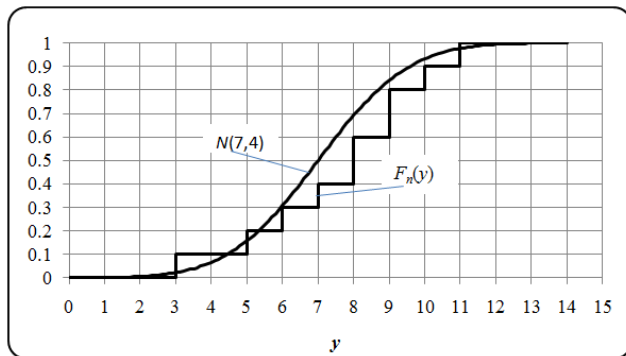


Figure 11.2. Hypothetical and empirical *cdfs*

11.8 Kolmogorov-Smirnov goodness-of-fit test

Kolmogorov-Smirnov test

The *Kolmogorov-Smirnov test* or *K-S test* is used to decide if a sample comes from a population with a specified *cdf* $F^*(y)$. The test is based on the difference between the empirical *cdf* $F_n(y)$ and $F^*(y)$. One advantage of the *K-S test* over the chi-square test is that it does not depend on arbitrary classification of the data.

K-S test

Also, for small sample sizes, the K - S test can always be carried out, whereas the chi-square test may not be appropriate for small samples.

The K - S test however does have its limitations. The test is designed to be used with continuous variables, and the associated cdf must be completely specified. Also, the test is quite insensitive to discrepancies between the empirical and the hypothesized $cdfs$ in the tail areas of the distributions.

- Technically, the K - S test could be used for testing distributions of ordinal variables. However, such cases are rather limited and are usually handled better with a chi-square test.

As with the chi-square test, the K - S test is usually performed with a two-sided alternative:

$$H_0: F(y) = F^*(y) \quad (11.10)$$

$$H_1: F(y) \neq F^*(y)$$

The K - S test is based on the maximum discrepancy between $F_n(y)$ and $F^*(y)$. The test statistic is:

$$D = \max_i |F^*(y_{(i)}) - F_n(y_{(i)})| \quad (11.11)$$

where $y_{(i)}$ is the i^{th} ordered observation as in Equation (11.8).

Because both $F^*(y)$ and $F_n(y)$ are nondecreasing functions, the maximum discrepancy between the two must be either just before or just after a step is made in $F_n(y)$. Accordingly, D can be calculated as:

$$D = \max_i \left(\left| F^*(y_{(i-1)}) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F^*(y_{(i)}) \right| \right) \quad (11.12)$$

The K - S test rejects the null hypothesis if D exceeds the critical value $d_q(n)$ for sample size n and $q = 1 - \alpha$. These values are given in Table T-13, “Quantiles, $d_q(n)$, of the Kolmogorov-Smirnov D statistic,” of the appendix. If H_0 is rejected, we conclude that the statistical evidence does not support the claim that the sample came from a distribution with the hypothesized $cdf F^*(y)$.

Example 11.6. Example 11.4 revisited. In Example 11.4, we used the chi-square test to test whether the sample came from a normal distribution $N(310, 100)$. We now test the same sample using the K - S test. This test is shown using an Excel spreadsheet with step-by-step instructions. Again, the null and the alternative hypotheses are:

$$H_0: F(y) = F^*(y)$$

$$H_1: F(y) \neq F^*(y)$$

Step-by-step intermediate calculations for this example are given for Excel, and the numerical values are displayed in Table 11.10.

Step 1. In columns A2..A31, enter the sequential numbers of the observations.

Step 2. Enter the data in Excel in cells B2..B31. If the data are not sorted already, sort the data in ascending order by clicking “Data” then “Sort A → Z” from the menu to indicate sorting in ascending order.

Step 3. Calculate sample size n in cell C37 (not shown) using
=COUNT(B1..B31).

Step 4. Calculate the height of the empirical distribution function for the smallest observation $y_{(1)}$ in cell C2, using
=FREQUENCY(B\$2:B\$31,B2)/\$C\$37. Copy cell C2 to cells C3 through C33.

Step 5. Use =NORMDIST(B4,310,10,1) in cell D2 to enter the cumulative $N(310,100)$ value distribution for $y_{(1)}$ in cell D2. (The last entry of 1 in this function indicates that the cumulative, rather than marginal, value is requested.) Copy cell D2 to cells D3 through D33.

Step 6. In cell E2, enter =ABS(D4 - A4/\$C\$37) to calculate $|F^*(y_{(1)}) - F_n(y_{(1)})|$. Copy cell E2 to cells E3 through E33.

Step 7. In cell F2, enter =ABS(D4 - (A4-1)/\$C\$37) to calculate $|F^*(y_i) - F_n(y_{(i-1)})|$. Copy cell F2 to cells F3 through F33.

➤ Since $i - 1 = 0$ when $i = 1$, $F_n(y_{(i-1)})$ is defined as .000 for cell F2.

From the last two columns thus created, we find their maximum value,
 $D = \text{MAX}(E4..E33)$, which is $D = 0.222$.

From Table T-13 of the appendix, the critical value $d_{0.95}(30) = 0.242$. Since D does not exceed the critical value, we have no statistical evidence to reject the $N(130, 100)$ hypothesis at the 5% level of significance. However, had we used a 10% level of significance, the critical value would have been $d_{0.90}(30) = 0.218$, and we would have rejected the null hypothesis. These are the same conclusions as drawn in Example 11.4. Neither the chi-square test nor the K - S test rejects H_0 at the $\alpha = 0.05$ level of significance, and both would have rejected H_0 if we had used $\alpha = 0.10$ instead.

In general, Conover (1980) suggests that the K - S test may be preferred over the chi-square test if the sample size is small, because exact critical values are readily available for the K - S test. Conover also states that the K - S test is more

powerful than the chi-square test for many situations. For further details and comparisons, see Slaktor (1965).

Table 11.10. Kolmogorov-Smirnov test applied to Example 11.4

	A	B	C	D	E	F
1	Rank, (i)	y	$F_n(y_{(i)})$	$F^*(y_{(i)})$	$ F^*(y_{(i)})-F_n(y_i) $	$ F^*(y_{(i)})-F_n(y_{(i-1)}) $
2	1	267	0.033	0.000	0.033	0.000
3	2	270	0.067	0.000	0.067	0.033
4	3	278	0.100	0.001	0.099	0.066
5	4	280	0.133	0.001	0.132	0.099
6	5	283	0.167	0.003	0.164	0.130
7	6	284	0.200	0.005	0.195	0.162
8	7	287	0.233	0.011	0.222	0.189
9	8	296	0.267	0.081	0.186	0.152
10	9	299	0.300	0.136	0.164	0.131
11	10	300	0.367	0.159	0.174	0.141
12	11	300	0.367	0.159	0.208	0.174
13	12	304	0.400	0.274	0.126	0.093
14	13	306	0.433	0.345	0.088	0.055
15	14	313	0.467	0.618	0.151	0.185
16	15	314	0.533	0.655	0.155	0.188
17	16	314	0.533	0.655	0.122	0.155
18	17	315	0.567	0.691	0.124	0.158
19	18	316	0.667	0.726	0.126	0.159
20	19	316	0.667	0.726	0.093	0.126
21	20	316	0.667	0.726	0.059	0.093
22	21	317	0.733	0.758	0.058	0.091
23	22	317	0.733	0.758	0.025	0.058
24	23	319	0.767	0.816	0.049	0.083
25	24	321	0.800	0.864	0.064	0.097
26	25	325	0.833	0.933	0.100	0.133
27	26	326	0.867	0.945	0.078	0.112
28	27	331	0.900	0.982	0.082	0.115
29	28	342	0.933	0.999	0.066	0.099
30	29	351	0.967	1.000	0.033	0.067
31	30	354	1.000	1.000	0.000	0.033

11.9 Shapiro-Wilk test for normality

Shapiro-Wilk test The *Shapiro-Wilk test*, also known as the *W-test*, tests whether the distribution from which a sample $\{y_1, y_2, \dots, y_n\}$ is taken is a normal distribution. The test requires a table of coefficients to compute the test statistic in addition to a table of critical values.

W-test

- The *W-test* is also applicable to distributions that can be related to the normal distribution by a transformation. For example, if we wish to test whether the observations came from a log-normal distribution, we take the logarithm of each observation and then apply the test to the transformed observations.

The *W-test* is applicable when neither μ nor σ is specified. The *W-test* is considered an omnibus test for normality because of its superiority to other procedures over a wide range of problems and conditions that depend on an assumption of normality. Based on comparison studies by Shapiro, Wilk, and Chen (1968) and LeBrecque (1977), the *W-test* appears to be superior to the chi-square test in many situations where n is no larger than 50. Its only limitation is that it is applicable only to sample sizes between 3 and 50.

The null and alternative hypotheses are:

$$\begin{aligned} H_0: & \text{The sample comes from a normal distribution.} \\ H_1: & \text{The underlying distribution is not normal.} \end{aligned} \quad (11.13)$$

In this section, we first present the general structure of the test and then illustrate the test in detail with an Excel spreadsheet. The calculations are lengthy but straightforward and are best carried out with a spreadsheet.

Let $\{y_{(1)}, y_{(2)}, \dots, y_{(n)}\}$ be a random sample of n independent observations arranged in ascending order. The test statistic is:

$$W = \frac{B^2}{(n-1)S^2} \quad (11.14)$$

where:

$$B = \sum_{i=1}^k a_i (y_{(n-i+1)} - y_{(i)})$$

$k = n/2$ if n is even or $(n - 1)/2$ if n is odd

$S^2 =$ sample variance

$a_i =$ coefficient obtained from Table T-6a, "Coefficients $\{a_{n-i+1}\}$ for the *W-test* for normality," in the appendix

The null hypothesis H_0 of normality is rejected at the α level of significance if w , the calculated value of W , is *less* than the critical value $w_q(n) = w_{1-\alpha}(n)$ obtained from Table T6-b, “Quantiles, $w_q(n)$, for the W -test for normality,” in the appendix. Note that in this table, the quantile $q = \alpha$. Also note that Tables T-6a and T-6b are limited to sample sizes between 3 and 50.

The W -test is illustrated in Example 11.7, taken from Bowen and Bennett (1988), p. 532). The procedure is detailed in 11 steps and is illustrated with an Excel spreadsheet.

Example 11.7. Testing normality with the W -test. Table 11.11 shows the percent uranium for 17 cans of ammonium diuranate (ADU) scrap. We wish to test whether the data come from a normal distribution.

Table 11.11. Percent uranium for ADU scrap

35.5	79.4	35.2	4.1	25.0	78.5
78.2	37.1	48.4	28.6	75.5	34.3
29.4	29.8	28.4	23.4	77.0	

Excel’s spreadsheet calculations are shown in Table 11.12 and the steps, keyed to the spreadsheet, are detailed along with that table.

The W -test is described here in an 11-step procedure.

- Step 1.** Arrange the n observations in *ascending* order, and enter the ordered data in cells B2..B18. Alternatively, enter the raw data in cells B2..B18, select those cells, and click “Data” and then “Sort A \rightarrow Z” to sort the data in ascending order.
- Step 2.** If n is even, set $k = n/2$; if n is odd, set $k = (n - 1)/2$. For Example 11.7, where $n = 17$, set $k = 8$.
- Step 3.** Rearrange the observations in *descending* order of magnitude, and enter the first k of them as shown in column C of Table 11.12. In Excel, you may first copy cells B2..B18 to cells C2..C18 and click “data” and then “Sort Z \rightarrow A” to sort the data in descending order. In Example 11.7, where $k = 8$, only 8 values are needed, and those are shown in cells C2..C9. To avoid confusion, delete cells C10..C18.

Table 11.12. W-test calculations for Example 11.7

	A	B	C	D	E	F
1	Rank (<i>i</i>)	Data, ascending order $y_{(i)}$	Data, descending order $y_{(n-i+1)}$	Difference, $y_{(n-i+1)} - y_{(i)}$	Table T-6a coefficients, a_i $n = 17, k = 8$	Cross product, $a_i(y_{(n-i+1)} - y_{(i)})$
2	1	23.4	79.4	56.0	0.4968	27.8208
3	2	25.0	78.5	53.5	0.3273	17.5106
4	3	28.4	78.2	49.8	0.2540	12.6492
5	4	28.6	77.0	48.4	0.1988	9.6219
6	5	29.4	75.5	46.1	0.1524	7.0256
7	6	29.8	48.4	18.6	0.1109	2.0627
8	7	34.3	40.1	5.8	0.0725	0.4205
9	8	35.2	37.1	1.9	0.0359	0.0682
10	9	35.5				
11	10	37.1			$B =$	77.1796
12	11	40.1				
13	12	48.4			$S^2 =$	476.5968
14	13	75.5			$n - 1 =$	16
15	14	77.0			$w =$	0.7811
16	15	78.2			$w_{.05}(17) =$	0.892
17	16	78.5				
18	17	79.4				

- Step 4.** Calculate the differences between the corresponding entries of column C and Column B of Table 11.12, and enter those differences in the corresponding D column. In Excel, enter = C2 - B2 in cell D2, and then copy cell D2 onto cells D3..D9.
- Step 5.** From Table T-6a of the appendix, copy the k coefficients $\{a_1, a_2, \dots, a_k\}$ associated with sample size n onto column E of Table 11.12. In Excel, these coefficients are entered in cells E2..E9.
- Step 6.** Multiply the associated elements of column D and column E onto column F of Table 11.12. In Excel, enter =D2*E2 in cell F2, and then copy cell F2 onto cells F3..F9.

Step 7. Sum the last column of Table 11.12. In Excel, enter =SUM(F2..F9) in cell F11. Denote the sum by B . In Example 11.7, the value of B is 77.1796.

Step 8. Calculate S^2 , the sample variance of the n observations. In Example 11.12, the sample variance is 476.5968. In Excel, enter =VAR(B2..B1) in cell F13.

Step 9. The test statistic, W , is computed from Equation (11.14) as:

$$w = \frac{77.1796^2}{(17-1)(476.5968)} = 0.7811$$

In Excel, enter $n - 1 = 16$ in cell F14, and calculate w in cell F15 as =F11^2/(F13*F14).

Step 10. From Table T-6b of the appendix, obtain the critical point $w_q(n) = w_{1-\alpha}(n)$ for the corresponding sample size and the appropriate level of significance. In Example 11.7, use $\alpha = 0.05$ and $n = 17$ to find $w_{0.05}(17) = 0.892$. This value is reported in cell F16 of the spreadsheet.

Step 11. Compare w from Step 9 to $w_q(n)$ in Step 10. If w is *smaller* than $w_q(n)$, the null hypothesis of normality is rejected. Because $0.781 < 0.892$, the null hypothesis in Example 11.7 is rejected. This indicates that the observations are unlikely to have come from a normal distribution.

It is important to note that, in practice, observations probably never come from a distribution that is exactly equal to a hypothesized distribution. In particular, samples rarely, if ever, come from a normal distribution. As a consequence, almost any goodness-of-fit test will result in rejection of the null hypothesis if the number of observations is very large. Thus, the outcome of such a test should be carefully interpreted. Goodness-of-fit tests provide a criterion for determining whether the agreement between the actual and hypothesized distributions is close enough that the hypothesized distribution provides a satisfactory approximation to the actual distribution. If the approximation is deemed satisfactory, then statistical methods based on the hypothesized distribution can be applied with some assurance that the results (inferences) are valid.

11.10 D'Agostino test for normality

D'Agostino test The ***D'Agostino test***, also known as the ***D' test***, tests whether the distribution from which a sample $\{y_1, y_2, \dots, y_n\}$ is drawn is a normal distribution. Like the W -test, the D' test is applicable when neither μ nor σ is specified.

- Just like the W -test, the D' test is also applicable to distributions that can be related to the normal distribution by a transformation. For example, if we wish to test whether the observations came from a log-normal

distribution, we take the logarithm of each observation and then apply the D' test to the transformed observations.

Just like the W -test, the D' test is considered an omnibus test for normality because of its superiority to other procedures over a wide range of problems and conditions that depend on an assumption of normality. The D' test complements the W -test, which is applicable only to samples no larger than 50, and can be used for any sample size greater than 50.

The test presented here is derived from D'Agostino's original paper (1971), pp. 341–348.

A protocol for the test may also be found in the American National Standards Institute (ANSI) N15.15-1974, p. 12.

The null and alternative hypotheses are:

H_0 : The random sample comes from a normal distribution.

H_1 : The underlying distribution is not normal.

In this section, we first present the general structure of the test and then illustrate the use of the test in detail with an Excel spreadsheet.

Let $\{y_{(1)}, y_{(2)}, \dots, y_{(n)}\}$ be a random sample of n independent observations arranged in ascending order. The test statistic is:

$$D' = \frac{T}{\sqrt{S^2(n-1)}} \quad (11.15)$$

where:

$$\begin{aligned} T &= (1 - (n+1)/2)y_{(1)} + (2 - (n+1)/2)y_{(2)} + \dots + (n - (n+1)/2)y_{(n)} \\ &= \sum_{i=1}^n (i - (n+1)/2)y_{(i)} \end{aligned} \quad (11.16)$$

S^2 = sample variance

- Be aware that the definition of S^2 used in the ANSI N15.15 standard differs from our definition of S^2 . Whereas we use S^2 to denote the sample variance, ANSI uses S^2 to denote the adjusted sum of squares, which is $(n-1)$ times the sample variance. Also, the ANSI test statistic is n^2 times the test statistic that is described below.

Because the values of the D' test have a small range for any given size sample, calculations should be carried out to at least five significant digits.

The D' test involves a comparison of the calculated D' value with two quantiles from Table T-14, “Quantiles, $D'_q(n)$, of the D' statistic,” of the appendix. The test is two-sided and requires two critical values that bound a noncritical region. For each combination of n and α , the critical values are found in Table T-14 under the row that corresponds to n and the columns for $q_{\alpha/2}(n)$ and $q_{1-\alpha/2}(n)$. As an example, for $n = 100$ and $\alpha = 0.05$, the critical values are 274.4 and 286.0. If the calculated D' is not between these two values, the null hypothesis is rejected.

Example 11.8. Testing normality with the D' test. Table 11.13 lists 68 measurements (in microcuries) in ascending order of surface area considered for clearance.

Table 11.13. Data for Example 11.8

3.15	4.48	8.27	10.52	10.52	11.09	11.2	11.2	11.2	13.5
15.1	15.2	15.3	15.7	15.89	16.1	17.66	19.85	20	20.2
20.2	20.2	20.4	20.4	20.4	20.4	22.4	22.6	22.98	23.27
25.1	25.2	27.2	27.2	27.5	27.7	29.3	29.7	33.7	35
35.5	38.1	39.5	44.2	44.2	46.71	50.84	52.96	53.8	55.65
63.3	70.1	71.6	74.5	74.9	81.1	90.9	149	163.84	204.7
251.3	306.6	311.83	327.42	356.66	445.15	462.21	502.97		

Before any further investigation, we wish to test whether the data are distributed normally. The test is conducted in the following steps. The steps are keyed to the Excel spreadsheet in Table 11.14. Note that the table is divided into two groups of columns to accommodate the calculations on one page.

- Step 1.** Enter the numbers 1, 2, ..., n onto the column labeled “Rank, i .” In the spreadsheet, these are entered in cells A2..A35, E2..E35.
- Step 2.** Enter the n sample observations in ascending order onto the column labeled $y_{(i)}$. In the spreadsheet, these are entered in cells B2..B35, F2..F35.
- Step 3.** Calculate, for each i , the quantity $i - (n+1)/2$, and enter that quantity onto the column so labeled. For $n = 68$, this entry would be $i - 34.5$. In the spreadsheet, these are entered in cells C2..C35, G2..G35. Thus, in cell C2, enter =A1-34.5. Copy this cell onto cells C3..C35, G2..G35.
- Step 4.** For each i , multiply $(i - (n+1)/2)$ by $y_{(i)}$, and enter the product in the column so labeled. In the spreadsheet, enter =B2*C2 in cell D2. Copy this cell onto cells D3..D35, H2..H35.

- Step 5.** Add the values in the columns labeled $(i - (n+1)/2)y_{(i)}$. Denote the calculated sum by t , and enter it in any empty cell. In the spreadsheet, t is calculated as =SUM(D2..D35, H2..H35). In this example, $t = (-105.53) + (-145.6) + \dots + 16849.495 = 111392.84$.
- Step 6.** Calculate the sample variance, S^2 . In Example 11.8, $S^2 = 13730.78$. In Excel, this is calculated as =VAR(B2..B35, F2..F35).
- Step 7.** Calculate D' using Equation (11.15). In Example 11.8, we calculate:

$$D' = \frac{111392.84}{\sqrt{(13730.78)(67)}} = 116.14$$

- Step 8.** Consult Table T-14 of the appendix. Under the row that corresponds to $n = 68$, obtain the critical points $D'_{0.025}(68) = 152.8$ and $D'_{0.975}(68) = 160.6$. Since D' does not lie between these two values, the null hypothesis is rejected.

Table 11.14. D' test calculations for Example 11.8

	A	B	C	D	E	F	G	H
1	Rank, i	$y_{(i)}$	$i-(n+1)/2$	$(i-(n+1)/2)y_{(i)}$	Rank, i	$y_{(i)}$	$i-(n+1)/2$	$(i-(n+1)/2)y_{(i)}$
2	1	3.15	-33.5	-105.53	35	27.5	.5	13.75
3	2	4.48	-32.5	-145.6	36	27.7	1.5	41.55
4	3	8.27	-31.5	-26.51	37	29.3	2.5	73.25
5	4	1.52	-3.5	-32.86	38	29.7	3.5	103.95
6	5	1.52	-29.5	-31.34	39	33.7	4.5	151.65
7	6	11.09	-28.5	-316.07	40	35	5.5	192.5
8	7	11.2	-27.5	-308	41	35.5	6.5	23.75
9	8	11.2	-26.5	-296.8	42	38.1	7.5	285.75
10	9	11.2	-25.5	-285.6	43	39.5	8.5	335.75
11	10	13.5	-24.5	-33.75	44	44.2	9.5	419.9
12	11	15.1	-23.5	-354.85	45	44.2	1.5	464.1
13	12	15.2	-22.5	-342	46	46.71	11.5	537.165
14	13	15.3	-21.5	-328.95	47	5.84	12.5	635.5
15	14	15.7	-2.5	-321.85	48	52.96	13.5	714.96
16	15	15.89	-19.5	-309.86	49	53.8	14.5	78.1
17	16	16.1	-18.5	-297.85	50	55.65	15.5	862.575
18	17	17.66	-17.5	-309.05	51	63.3	16.5	1044.45
19	18	19.85	-16.5	-327.53	52	7.1	17.5	1226.75
20	19	20	-15.5	-310	53	71.6	18.5	1324.6
21	20	2.2	-14.5	-292.9	54	74.5	19.5	1452.75
22	21	2.2	-13.5	-272.7	55	74.9	2.5	1535.45
23	22	2.2	-12.5	-252.5	56	81.1	21.5	1743.65
24	23	2.4	-11.5	-234.6	57	9.9	22.5	2045.25
25	24	2.4	-1.5	-214.2	58	149	23.5	3501.5
26	25	2.4	-9.5	-193.8	59	163.84	24.5	4014.08
27	26	2.4	-8.5	-173.4	60	204.7	25.5	5219.85
28	27	22.4	-7.5	-168	61	251.3	26.5	6659.45
29	28	22.6	-6.5	-146.9	62	306.6	27.5	8431.5
30	29	22.98	-5.5	-126.39	63	311.83	28.5	8887.155
31	30	23.27	-4.5	-104.72	64	327.42	29.5	9658.89
32	31	25.1	-3.5	-87.85	65	356.66	3.5	10878.13
33	32	25.2	-2.5	-63	66	445.15	31.5	14022.225
34	33	27.2	-1.5	-4.8	67	462.21	32.5	15021.825
35	34	27.2	-0.5	-13.6	68	502.97	33.5	16849.495

12

Contingency tables

12.1 What to look for in Chapter 12

Chapter 12 brings us to contingency tables and their analyses. Such analyses have many applications in experiments involving enumeration data. Special topics and ideas include:

- *general structure of contingency tables, §12.3*
- *independence and interaction, §12.4, §12.5*
- *chi-square statistic, §12.5*
- *$r \times c$ contingency tables, §12.5*
- *shortcut formula for 2×2 contingency tables, §12.6*
- *Fisher's exact probability test, §12.7*
- *Simpson's paradox, §12.8*
- *McNemar's test statistic, §12.9*

12.2 Contingency tables

contingency table **Contingency table** analysis is used for enumeration data to test whether two data classification criteria, such as gender and performance, are related. The main test for investigating such relations is a chi-square test. This test is very similar to the chi-square test for goodness-of-fit described in Chapter 11.

This chapter breaks a long-standing tradition honored in many books on applied statistics where the treatment of contingency table analysis is positioned near the end of the text—if the topic is discussed at all. In this chapter, you will find material on the construction, analysis, and interpretation of contingency tables that is of special value in the development of this text for the following reasons:

- Contingency tables can be described and illustrated with very little, and usually uncomplicated, data.
- Contingency tables can be used to illustrate many of the concepts required in making statistical inferences without the need for complex calculations.
- Contingency tables arise naturally in many data-driven situations, their analysis is straightforward, and they provide a valuable tool that can be used in dealing with daily problems.
- Contingency tables can illustrate situations where statistical naiveté can lead to incorrect inferences.

12.3 Structure of contingency tables

factor Contingency tables deal with counting data, where observations are classified by two qualitative variables, called **factors**. For example, one factor may be education and the other may be company position. Each factor has two or more **levels** or **categories**. For example, the levels of the education factor may be high school, college, and graduate school, while the levels of the company position factor may be management and nonmanagement. The goal of contingency table analysis is to investigate if there is any relation between the factors (e.g., whether education and company position are related).

level

category

The term “related” does not imply cause and effect. It does, however, imply mutual dependency, or “contingency,” where the level of one factor may be related to the level of the other factor. Regardless of the degree of relatedness found by any contingency table analysis, the question of whether there is any causal relation between the factors is beyond this analysis. This question is not a statistical one and can be investigated only with a separate subject-specific analysis.

Here is one definition, paraphrasing Kendall and Buckland (1971, p. 32):

A **contingency table** is an array of observations set out in a two-way table, the rows and the columns of which are any two qualitative characteristics (i.e., factors) possessed by the items, provided that no row or column is completely empty. For example, if the first characteristic *A* is *r*-fold (i.e., *r* levels) and the second characteristic *B* is *c*-fold (i.e., *c* levels), then the resulting display is said to be an *r* × *c* contingency table; that is, it is a two-way table with *r* rows and *c* columns.

Table 12.1 shows a general *r* × *c* contingency table, where the unshaded area shows the count for each of the *r* × *c* cells. The count of observations in cell (*i*, *j*) is denoted by *O_{ij}*. By convention, the first subscript, *i*, denotes the row, and the second subscript, *j*, denotes the column. The totals shown in the last row and the last column of Table 12.1 are the **marginal totals**. The sum of the marginal totals (for either the rows or the columns) is the **grand total** and is typically denoted by *n*.

marginal totals
grand total

Table 12.1. Structure of a contingency table

		Factor 2: Columns				Total
		1	2	...	<i>c</i>	
Factor 1: Rows	1	<i>O₁₁</i>	<i>O₁₂</i>	...	<i>O_{1c}</i>	$O_{1+} = \sum_{j=1}^c O_{1j}$
		<i>O₂₁</i>	<i>O₂₂</i>	...	<i>O_{2c}</i>	$O_{2+} = \sum_{j=1}^c O_{2j}$
	⋮	⋮	⋮	⋮	⋮	⋮
	<i>r</i>	<i>O_{r1}</i>	<i>O_{r2}</i>	...	<i>O_{rc}</i>	$O_{r+} = \sum_{j=1}^c O_{rj}$
Total	$O_{+1} = \sum_{i=1}^r O_{i1}$	$O_{+2} = \sum_{i=1}^r O_{i2}$...	$O_{+c} = \sum_{i=1}^r O_{ic}$	Grand total = O_{++} $= \sum_{i=1}^r O_{i+} = \sum_{j=1}^c O_{+j}$ $= \sum_{i=1}^r \sum_{j=1}^c O_{ij} = n$	

In this chapter, *O_{ij}* will always denote the actual count in cell (*i*, *j*), i.e., *O_{ij}* denotes a number and not a random variable. This notation is consistent with the standard notation for contingency tables.

plus notation The notation for the marginal and grand totals follows a convention, called the **plus notation**, where the plus in the subscript indicates the sum over the index of the “plussed” subscript. Thus, O_{2+} is the sum of all O_{ij} observations whose first subscript is 2, and O_{+3} is the sum of all O_{ij} observations whose second subscript is 3. Following the same notation, the sum n of all observations, may be written as O_{++} .

dot notation Note that many authors use the **dot notation** (such as $O_{2.}$, $O_{.3}$, or $O_{..}$) to denote the various sums. We use the plus notation to remind us of the summation operation involved and to avoid confusion when using a dot at the end of a sentence.

Example 12.1. Computer literacy and age. A computer literacy test was given to 80 men between the ages of 20 and 50. The results are summarized in Table 12.2, where the rows represent three age groups, and the columns represent passing or failing grades on the test. The natural question that one may ask is whether passing the test depends on one’s age. Because this question implies a causal relation, which contingency table analysis does not address, a more appropriate question is whether age and test grade are related.

Table 12.2. A 3×2 contingency table for Example 12.1

	Pass	Fail	Row totals
Age: 20–29	24	6	$O_{1+} = 30$
Age: 30–39	16	14	$O_{2+} = 30$
Age: 40–50	7	13	$O_{3+} = 20$
Column totals	$O_{+1} = 47$	$O_{+2} = 33$	$n = O_{++} = 80$

Section 12.5 presents the analysis of the data in Example 12.1.

12.4 Independence

Suppose for a moment that Table 12.1 represents an entire population of n items. If an item is selected at random from this population, the probability that that item would be in cell (i,j) is $p_{ij} = O_{ij}/n$. Similarly, the probability that it would be in row i is $p_{i+} = O_{i+}/n$, and the probability that it would be in column j is $p_{+j} = O_{+j}/n$.

Suppose further that the rows and columns are independent (i.e., the two classification factors are independent). Then, by Equation (6.18), the following equation holds:

$$p_{ij} = (p_{i+})(p_{+j}), \quad i = 1, \dots, r, \quad j = 1, \dots, c \quad (12.1)$$

interaction The absence of independence is called **interaction**. If the rows and columns are not independent, they are said to interact or be interacting. For interacting factors, Equation (12.1) does not hold.

Equation (12.1) may be regarded as a null hypothesis in testing whether the two classification factors (rows and columns) are independent or interact. The null and alternative hypotheses are:

H_0 : Rows and columns are independent.

H_1 : Rows and columns interact.

(12.2)

12.5 Testing independence

Now suppose Table 12.1 presents the results of a random sample of size n from a population. For this population, p_{ij} is estimated by O_{ij}/n , p_{i+} is estimated by O_{i+}/n , and p_{+j} is estimated by O_{+j}/n . Suppose for a moment that these estimates of p_{i+} and p_{+j} are exact. If the rows and the columns are independent, from Equation (12.1), the expected value of the random variable whose observed value is O_{ij}/n is:

$$\frac{E_{ij}}{n} = \left(\frac{O_{i+}}{n} \right) \left(\frac{O_{+j}}{n} \right) \quad (12.3)$$

expected count Under these assumptions, the **expected count** E_{ij} for cell (i,j) is:

$$E_{ij} = \frac{(O_{i+})(O_{+j})}{n} \quad (12.4)$$

expected frequency E_{ij} is also called the **expected frequency**. Note that E_{ij} does not have to be (and very rarely is) an integer.

Note from Equation (12.3) that the expected count for any cell is the cross-product of the marginal (row and column) totals of that cell divided by the grand total. For example, the expected count for cell (1,1) of Table 12.2 is $(30)(47)/80 = 17.625$.

Only on very rare occasions do the observed counts exactly agree with the expected counts. Failure to agree may be attributed to one or both of two reasons:

- There are random fluctuations.
- Rows and columns are not independent.

To summarize, for every cell in the $r \times c$ table, we have an observed count O_{ij} and, under the assumption of independence, an expected count E_{ij} . The differences $O_{ij} - E_{ij}$ are the key elements in testing the null hypothesis of independence.

The test statistic for independence, developed by Karl Pearson (1904), is given by the chi-square statistic, χ^2 , where:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (12.5)$$

- Note the similarity of this chi-square statistic to the chi-square statistic from Equation (11.2) used for the chi-square goodness-of-fit test.

A mathematically equivalent form of Equation (12.5) is given by Equation (12.6):

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed count})^2}{\text{expected count}} - n = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij})^2}{E_{ij}} - n \quad (12.6)$$

The chi-square statistic has associated degrees of freedom, denoted by ν . For an $r \times c$ contingency table, the degrees of freedom is:

$$\nu = (r - 1)(c - 1) \quad (12.7)$$

The chi-square statistic is typically written with the degrees of freedom in parentheses, as in $\chi^2(\nu)$.

- The chi-square test is a one-sided test. Only values larger than the critical value lead to the rejection of H_0 .
- Be aware that although the statistic is called chi-square, its distribution only approximates the χ^2 distribution. However, this approximation is excellent even for small samples, provided that the expected count of each cell is not too small, as discussed in Section 11.4.

Applying Equation (12.5) or (12.6) to the contingency table, we obtain the calculated chi-square, $\chi^2(\nu)$. This is used to test H_0 in Equation (12.2).

Returning to Example 12.1, the calculations for Equations (12.5) and (12.6) are shown in Table 12.3. Note that the expected count exceeds 8 in all cells, thus satisfying the criterion in Section 11.4.

Table 12.3. Calculations for the 3×2 contingency table of Example 12.1

Row	Column	Observed, O	Expected, E	$(O - E)$	$(O - E)^2$	$(O - E)^2/E$	O^2/E
1	1	24	17.625	6.375	40.641	2.306	32.681
1	2	6	12.375	-6.375	40.641	3.284	2.909
2	1	16	17.625	-1.625	2.641	0.150	14.525
2	2	14	12.375	1.625	2.641	0.213	15.838
3	1	7	11.750	-4.750	22.562	1.920	4.170
3	2	13	8.250	4.750	22.562	2.735	20.485
Total		80	80.000			10.608	90.608
Calculated chi-square:		Using Equation (12.5), $\chi^2(2) = 10.608$					
$\nu = (3 - 1)(2 - 1) = 2$		Using Equation (12.6), $\chi^2(2) = 90.608 - 80 = 10.608$					

- Note that if Equation (12.5) is used, the column labeled O^2/E is superfluous. If Equation (12.6) is used, the columns labeled $(O - E)$, $(O - E)^2$, and $(O - E)^2/E$ are superfluous. This demonstrates that Equation (12.6) is easier to use, especially for hand calculation.

Next, the calculated $\chi^2(2)$ is compared to a critical value to determine statistical significance. For $\alpha = 0.05$ and $\nu = 2$, we consult Table T-2 of the appendix under the quantile $q = 1 - \alpha = 0.95$ to find the value of 5.99. Since $\chi^2(2) = 10.608$ is greater than 5.99, we reject H_0 . The statistical evidence indicates that participant age and passing the literacy test are not independent.

- There may be more to this story. Other factors may affect passing of the test, such as education, gender, and related experience.

Any one of the following statements could summarize our findings:

- The statistical test finds that age and test performance are not independent ($p < 0.05$).
- We conclude that age and test performance are dependent ($p < 0.05$).
- Age and test performance interact at the 0.05 level of significance.

If $\chi^2(2)$ were not larger than 5.99, our findings could be summarized as follows:

- There is no statistical evidence that age and test performance are dependent.
- Age and test performance do not appear to be dependent.

12.6 Special case: 2x2 contingency tables

The 2x2 contingency table warrants special attention as such a table can represent many situations. We start with Equation (12.8) which, for a 2x2 contingency table, is mathematically equivalent to Equations (12.5) and (12.6). Equation (12.8) is a special shortcut formula for 2x2 contingency tables only. There is no equivalent formula for larger contingency tables.

$$\chi^2(1) = \frac{(n)[(O_{1+})(O_{+2}) - (O_{2+})(O_{+1})]^2}{(O_{1+})(O_{2+})(O_{+1})(O_{+2})} \quad (12.8)$$

From Equation (12.8), we can see the pattern in using the shortcut formula.

- The term in brackets is the difference between the products of the diagonal cells.
- The numerator is the square of the term in brackets multiplied by the grand sum, n .
- The denominator is the product of all the marginal sums.

Example 12.2. Weld inspection. Two teams, A and B, perform a welding operation. A regional inspector periodically evaluates their welds for acceptability. Table 12.4 summarizes the inspection results for 100 welds, selected at random from a single day's work.

Table 12.4. A 2x2 contingency table for weld inspections

	Column 1: Acceptable	Column 2: Not Acceptable	Row totals
Row 1: Team A	30	0	30
Row 2: Team B	50	20	70
Column totals	80	20	100

We note that each data cell has an expected count of at least 6, thus satisfying the criterion of Section 11.4.

Using Equation (12.8), the chi-square statistic is calculated as:

$$\chi^2(1) = \frac{(100)[(30)(20) - (50)(0)]^2}{(80)(20)(30)(70)} = 10.71$$

The critical value from Table T-2 is $\chi_{0.95}^2(1)$, or 3.84. Since the critical value is exceeded, we conclude that team affiliation and weld quality interact (i.e., they are dependent).

Example 12.3. Mastering technical material. We examine whether Internet support is related to students' mastery of technical material. A class of 170 students was classified into four groups as shown in Table 12.5. The expected count for each cell is sufficient for running a chi-square test.

Table 12.5. A 2×2 contingency table for the mastery study

	Mastered	Did not master	Row totals
Internet	46	14	60
No internet	19	91	110
Column totals	65	105	170

The calculated chi-square statistic is:

$$\chi^2(1) = \frac{170[(46)(91) - (19)(14)]^2}{(60)(110)(65)(105)} = 58.0$$

The critical value is $\chi_{0.95}^2(1) = 3.84$. Since χ^2 exceeded the critical value, we conclude that mastering the material and having Internet support are dependent. Note that we cannot conclude that having Internet support helped the students master the material. Another possible explanation is that the better students were more likely to get Internet support and would have mastered the material even without it.

Example 12.4. Age discrimination. In a large agency, 76 secretaries at a low grade level were promoted during a given calendar year. In the same year, 213 secretaries at the same grade level were not promoted. All $76 + 213 = 289$ secretaries were classified according to whether they had reached age 40 when the promotions were made. Table 12.6 presents the data.

Table 12.6. A 2×2 contingency table for testing age discrimination

	Promoted	Not promoted	Row totals
Age <40	44	104	148
Age ≥ 40	32	109	141
Column totals	76	213	289

The calculated chi-square is:

$$\chi^2(1) = \frac{289[(44)(109) - (104)(32)]^2}{(148)(141)(76)(213)} = 1.84$$

The critical value is $\chi_{0.95}^2(1) = 3.84$. Since the test statistic does not exceed the critical value, we conclude that there is no statistical evidence of age discrimination for secretaries at the grade level we investigated.

In this study, we note the following:

- We don't have a random sample here because everybody of concern is included. However, the data in Table 12.6 may be regarded as a sample of the agency's practice over the years.
- This study is restricted to a single occupation at a single grade level and does not necessarily reflect promotion practices at other grade levels or occupations.
- The test does not account for other possible factors, such as performance, education, or time in grade.

In our treatment of contingency tables, we address only two-factor data classification. The extension of contingency tables to k dimensions is not a trivial matter. Until the general availability of substantial and inexpensive computing resources in the 1970s and 1980s, researchers were limited to particular extensions of the two-dimensional contingency table to only three or four dimensions. However, methods for dealing with k -dimensional tables have been developed. The interested reader is referred to Rosner (2010), Section 13; Agresti (1990); and Upton (1978).

12.7 Fisher's exact probability test

Sample size considerations that apply to the chi-square goodness-of-fit test also apply to the chi-square test used in a contingency table analysis. Section 11.4 gives two rules for sample size stated in terms of the expected count in each cell. One rule suggests that in none of the cells should the expected count be smaller than 5. This rule is especially applicable to a 2×2 table.

One way to satisfy a minimum expected count rule is to reduce the number of rows or columns so as to increase the expected count in the deficient cells. Such a reduction, however, cannot be done in a 2×2 contingency table. Failing that, we resort to **Fisher's exact probability test** (Fisher (1970), Section 21.02), also known as **Fisher's exact test**, for independence in a 2×2 contingency table. This test is based on the hypergeometric distribution introduced in Section 8.6.

Using the notation of Table 12.1, a general 2×2 contingency table is shown in Table 12.7.

Table 12.7. A general 2×2 contingency table

	Column 1	Column 2	Row totals
Row 1	O_{11}	O_{12}	O_{1+}
Row 2	O_{21}	O_{22}	O_{2+}
Column totals	O_{+1}	O_{+2}	$O_{++} = n$

Let $\pi_{o_{11}, o_{12}, o_{21}, o_{22}}$ denote the probability of observing the specific values $\{o_{11}, o_{12}, o_{21}, o_{22}\}$ of the random variable when the marginal totals are fixed. From Equation (8.14), this probability is:

$$\pi_{o_{11}, o_{12}, o_{21}, o_{22}} = \frac{(o_{1+}!) (o_{2+}!) (o_{+1}!) (o_{+2}!)}{n! (o_{11}!) (o_{12}!) (o_{21}!) (o_{22}!)} \quad (12.9)$$

Fisher's exact test considers all possible table configurations that would yield the same marginal totals as those of the observed data. Using Equation (12.9), the probability of the observed configuration and all possible configurations that are *more* inconsistent with H_0 than the observed configuration are summed. If that sum is smaller than the level of significance α , we reject H_0 and conclude that the statistical evidence shows that the two factors (row and columns) are dependent. We illustrate this test in Example 12.5.

Example 12.5. Paint adhesion. A laboratory tests two paints for adhesion to concrete surfaces. A scratch test is performed on 10 surfaces painted with Paint A and on 18 surfaces painted with Paint B, where all surfaces are of equal size. Table 12.8 shows the results (pass/fail) of the tests. The null hypothesis is that passing the test and the type of paint are independent.

➤ An equivalent way to word the hypothesis of independence is:

H_0 : The proportion of Paint A surfaces that passes the test is the same as that of Paint B surfaces.

The topic of equality of proportions is considered in Chapter 22, "Binomial Experiments."

Table 12.8. Testing paint adhesion

	Pass	Fail	Totals
Paint A	2	8	10
Paint B	13	5	18
Totals	15	13	28

Using Equation (12.9), we calculate the probability of obtaining the specific values of 2, 8, 13, and 5 when the marginal totals are as shown in Table 12.8. This probability is:

$$\pi_{2,8,13,5} = \frac{10!18!15!13!}{28!2!8!13!5!} = 0.01030$$

Under the assumption of independence, the expected count in cell (1,1) is $(15)(10)/28 = 5.4$. Hence, the only counts in cell (1,1) that would be more inconsistent with H_0 than the observed count of 2 would be counts of 1 and 0. Tables 12.8a and 12.8b give the configurations and probabilities for these hypothetical counts.

- Note that all the other cell counts are determined once the cell (1,1) count is given. For example, in Table 12.8a, the cell (1,2) count is $10 - 1 = 9$, the cell (2,1) count is $15 - 1 = 14$, and the cell (2,2) count is $18 - 14 = 13 - 9 = 4$.
- In this example, we chose cell (1,1) as the “pivot cell” for our calculations because there are fewer additional configurations to be evaluated if the pivot cell has the smallest count. If two or more cells have the same smallest count, then any one may be chosen as the pivot cell.

Table 12.8a. Testing paint adhesion, cell (1,1) count = 1

	Pass	Fail	Totals
Paint A	1	9	10
Paint B	14	4	18
Totals	15	13	28

$$\pi_{1,9,14,4} = \frac{10!18!15!13!}{28!1!9!14!4!} = 0.00082$$

Table 12.8b. Testing paint adhesion, cell (1,1) count = 0

	Pass	Fail	Totals
Paint A	0	10	10
Paint B	15	3	18
Totals	15	13	28

$$\pi_{0,10,15,3} = \frac{10!18!15!13!}{28!0!10!15!3!} = 0.00002$$

The sum of the probabilities of the observed and more extreme configurations (i.e., configurations more inconsistent with H_0) is $0.01030 + 0.00082 + 0.00002 = 0.011$. Since this probability is smaller than $\alpha = 0.05$, we reject H_0 and conclude that passing the test and the type of paint are dependent.

Most statistical packages provide a routine for Fisher's exact probability test. Some of these packages provide this routine by default whenever a chi-square test for a contingency table is requested, while other statistical packages provide this routine by command. Check the software manual for instructions.

12.8 Simpson's paradox—better watch out!

Simpson's paradox

Simpson's paradox refers to the reversal in conclusions when data in groups are combined to form a single group. This paradox is illustrated by a celebrated example taken from *Science* (Bickel et al. (1975)) in which female students claimed discrimination in admission to one of the school's graduate programs. The investigation showed, however, that in virtually every department, females had a *higher* admission rate than that of their male counterparts.

We simplify the problem by using hypothetical numbers for a fictitious university consisting of only two schools.

Example 12.6. Admission records in two schools. Each of the two schools in a university processes 600 applications for a total of 1,200 applications. Table 12.9 gives the admit/deny versus male/female records for both schools. We test the null hypothesis of independence at the 5% level of significance.

Table 12.9. Admission records aggregated for two schools

	Admit	Deny	Totals
Male	250	300	550
Female	250	400	650
Totals	500	700	1200
$\chi^2(1) = 5.99$			

The chi-square test statistic is calculated from Equation (12.8) as $\chi^{2(1)} = 5.99$. This exceeds the critical value $\chi_1^2 = 3.84$, thus rejecting H_0 . We conclude that gender and admission in the university are dependent and claim that this constitutes statistical evidence of possible gender discrimination.

Next, the same records are tabulated for each school separately. Table 12.9a presents the records for School A.

Table 12.9a. Admission records for School A

	Admit	Deny	Totals
Male	200	200	400
Female	100	100	200
Totals	300	300	600
$\chi^2(1) = 0.00$			

Table 12.9b presents the records for School B:

Table 12.9b. Admission records for School B

	Admit	Deny	Totals
Male	50	100	150
Female	150	300	450
Totals	200	400	600
$\chi^2(1) = 0.00$			

The chi-square statistics for both schools are exactly zero. Thus, there is absolutely no statistical evidence of gender discrimination in admission to either school.

The paradox arises if we aggregate the data for both schools and conclude that, for the university as a whole, there is statistical evidence of possible gender discrimination. The resolution to the paradox is the realization that combining the two schools ignores the fact that different schools attract students with different interests and have different admission policies and admission rates.

Obviously, in the example cited, combining the individual contingency tables is inappropriate. In practice, when the data are broken into groups with multiple contingency tables, we should perform the analysis both ways: individual and aggregate. If the results all agree, it is appropriate to present the aggregate analysis as a summary. Otherwise, we should present each individual analysis and state that the individual groups should not be combined.

12.9 A contingency table look-alike: McNemar's test statistic

McNemar's test

McNemar's test (McNemar (1947)) is designed for situations where it is planned to treat the items in a population to enhance their performance or quality. However, before applying the treatment, we wish to test whether the treatment has any effect.

sample items are treated and are then retested. We evaluate the treatment by comparing the test results before and after treatment. In this situation, all items in the sample serve as their own before/after controls.

The data structure for McNemar's test bears a strong resemblance to the usual contingency table data structure discussed earlier in this chapter. However, although the data are assembled in a 2×2 table, it is not a contingency table as defined in Section 12.2, because the rows and columns do not correspond to factors. Consequently, the purpose, calculation, and interpretation of McNemar's test are quite different from those of the chi-square test for contingency tables.

Table 12.10 shows the data assembled in a 2×2 table. The cell counts are denoted by a , b , c , and d , as indicated. McNemar's test statistic employs the sensible principle that only those cases that show a change from one state to the next (fail followed by pass or pass followed by fail) are relevant. In the table, these are the circled cells whose counts are given by b and c .

Table 12.10. 2×2 table for McNemar's test statistic

	State 2		
State 1	Pass	Fail	Totals
Pass	a	b	$a + b$
Fail	c	d	$c + d$
Totals	$a + c$	$b + d$	$a + b + c + d$

The test statistic is given by Equation (12.10). Under the null hypothesis of no treatment effect, χ^2 is approximately distributed as a chi-square variable with $\nu = 1$.

$$\chi^2 = \frac{b - c}{b + c}^2 \quad (12.10)$$

If $b = c$, the treatment obviously has no effect, and no statistical test is needed.

The literature contains two variations of Equation (12.10), and each is also approximately distributed as a chi-square variable. These variations are given by Equations (12.11) and (12.12).

$$\chi^2 = \frac{|b - c| - 0.5}{b + c}^2 \quad (12.11)$$

$$\chi^2 = \frac{|b - c| - 1}{b + c}^2 \quad (12.12)$$

We use Equation (12.10) because it is more commonly used than either variation.

We illustrate the operation of McNemar's test with an example.

Example 12.7. Effect of laundering on protective mask filters.

A manufacturer wants to test the effect of laundering on a mask filter used to protect inspectors who enter a contaminated area. Each filter in a sample of 70 filters is first exposed to the same contaminated environment. A chemical test is then conducted to determine the acceptability (on a pass/fail basis) of each filter by comparing the contaminant's contents on the two surfaces (i.e., the exposed and the protected surfaces) of the filter. The filters are then laundered, and the exposure and the chemical test protocols are repeated. Some filters pass both pre- and post-laundry tests, others fail both, and still others pass one test and not the other. Table 12.11 shows the results of this investigation.

Table 12.11. Laundering effect on protective mask filters

State 1: Before laundry	State 2: After laundry		Totals
	Pass	Fail	
Pass	24	16	40
Fail	13	17	30
Totals	37	33	70

Using Equation (12.10), we obtain:

$$\chi^2 = \frac{(|16 - 13|)^2}{16 + 13} = 0.310$$

We test the null hypothesis at $\alpha = 0.05$. The calculated test statistic does not exceed the critical value $\chi^2_{0.95}(1) = 3.84$. We conclude that there is no significant statistical evidence of a laundering effect.

13

Tests of statistical hypotheses: One mean

13.1 What to look for in Chapter 13

Chapter 13 formalizes the process of statistical inference about the mean of a normal distribution. This chapter is the first chapter that tests hypotheses about a population mean; Chapter 15 tests the equality of two means, and Chapter 16 tests the equality of several means. The presentation identifies the circumstances when such inferences are valid and points out issues of concern. Special terms and concepts, both new and revisited, are:

- *model*, §13.2
- *power of a test*, §13.4
- *power curve*, §13.4
- *operating characteristic (OC) curve*, §13.5
- *sample size for testing a mean when σ is known*, §13.9
- *sample size for testing a mean when σ is not known*, §13.10

This chapter provides a general procedure for testing a hypothesis about the mean of a normal distribution. The procedure has two versions:

- when σ is known, §13.2
- when σ is not known, §13.10

For each version, we consider both one-sided and two-sided alternatives.

This chapter is also applicable to samples from a nonnormal distribution when the distribution of the sample mean \bar{Y} is approximately normal. When this requirement fails, we can use a nonparametric test, as described in Chapter 25.

13.2 A test of the mean: σ known

We begin by revisiting an earlier example.

Example 13.1. Response time to alarm. We return to Example 10.2, where we investigated the claim by the Director of Security that the system's mean response time to an intrusion is no more than 10 seconds by testing the null hypothesis that $\mu = 10$ seconds against a right-sided alternative. For this example, we assume that the standard deviation is $\sigma = 3$ seconds. Based on a sample of $n = 100$ responses to random intrusions, we calculate a mean response time of $\bar{y} = 10.4$ seconds. The assumptions and facts are:

- The intrusions occur at random times and the response times are mutually independent.
- The response times come from a normal distribution. (Chapter 11 presents the tools to test such an assumption.)
- The standard deviation is $\sigma = 3$ ($\sigma^2 = 9$). (In Chapter 14, we will learn how to test assumptions about variances or standard deviations.)
- The sample size is $n = 100$.
- The null hypothesis is $H_0: \mu = \mu_0 = 10$. Under H_0 , we write μ with a zero subscript to emphasize that this is the hypothesized mean.
- The alternative hypothesis is $H_1: \mu > 10$. H_1 is a right-sided alternative since we are concerned about an excessive average response time. However, we shall reconsider the alternative hypothesis in section 13.3.
- The level of significance is set at the default value of $\alpha = 0.05$.

model The response Y_i can be written as a *model*. In this context, a model is a mathematical expression of the response as a function of the parameters and assumptions. Specifically,

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \dots, n \quad (13.1)$$

where ε_i denotes the error (departure from the mean) of the i^{th} observation and

$$\varepsilon_i \sim N(0, \sigma^2) \quad (13.2)$$

The sample mean \bar{Y} is normally distributed with a mean μ and standard deviation σ/\sqrt{n} . The test statistic used for testing H_0 is the standard normal statistic:

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} = (\bar{Y} - \mu_0)\sqrt{n}/\sigma \quad (13.3)$$

From Equation (13.3), we see that the Z statistic comprises four components, and the statistical test essentially examines how these components fit together. These components are:

- what we know (here, $n = 100$)
- what we assume (here, $\sigma = 3$)
- what the data show (here, $\bar{y} = 10.4$)
- what is hypothesized (here, $H_0: \mu = \mu_0 = 10$)

Substituting these values into Equation (13.3), we get:

$$z = \frac{10.4 - 10.0}{3/\sqrt{100}} = 1.333$$

From Table T-1 of the appendix, the critical value is $z_{0.95} = 1.645$. Since $z = 1.333$ is not larger than the critical value, we do not reject H_0 . We conclude that there is no statistical evidence that the mean response time μ is larger than 10 seconds.

13.3 A one-sided test: A different view

Example 13.1 focuses on testing a claim by Security that the mean response time μ is no more than 10 seconds. Although not explicitly stated in the example, there may well be a second party involved—a regulator who does not want μ to be too large, and who expresses this concern by the right-sided alternative hypothesis $H_1: \mu > 10$. However, the test of hypothesis in Example 13.1 is structured to protect Security's interest but not necessarily a regulator's interest. To see why this is so, we show in Section 13.4 that H_0 would not have been rejected even for \bar{y} as large as 10.494. Recall that the mean of \bar{Y} is μ and that there is a 50% probability that \bar{Y} will be greater than its mean. Then, if $\mu = 10.494$, the probability of rejecting H_0 would be only 0.50. From a

regulator's perspective, it might be unacceptable to have such a low probability of rejection when the value of μ is almost half a second larger than the maximum average response time claimed by Security.

The answer to this dilemma is suggested in Section 10.6. If a regulator wishes to ensure a mean of no more than 10 seconds, H_1 needs to take a different direction. Specifically, we must use a left-sided alternative hypothesis written as $H_1: \mu < 10$. With this switch, the test is now structured to protect the regulator's interest. In the original setup in Example 13.1, Security's claim that $\mu \leq 10$ is supported if H_0 is not rejected. With the switched setup, the regulator's interest that μ not be too large is supported only if H_0 is rejected.

The test statistic is still Z from Equation (13.3), but now H_0 is rejected if z is *less* than the critical value. From Table T-1 of the appendix, the critical value is now $z_{0.05} = -1.645$. Clearly, H_0 is not rejected for any positive value of z , so the result of the switched test of hypothesis is the same as that for the original test in Example 13.1. This time, however, we conclude that the statistical evidence does not rule out the possibility that $\mu > \mu_0 = 10$. (Because the observed $\bar{y} > \mu_0$, this conclusion could have been reached without actually performing the test of hypothesis.) In fact, Equation (13.3) implies that H_0 would not have been rejected even for \bar{y} as small as 9.507, which would have yielded $z = -1.645$. Thus, with the switched test of hypothesis, \bar{y} can be no larger than 9.507 to reject H_0 and conclude that the statistical evidence supports Security's claim that $\mu \leq 10$.

The following is another example of the use of a left-sided alternative hypothesis.

*inventory
difference
(ID)*

Example 13.2. Inventory differences. A facility that produces special nuclear material (SNM) conducts a monthly inventory of its SNM and reports any significant shortage. For every item of SNM produced, the *inventory difference (ID)* is defined as the difference between the physical weight and recorded (book) weight of the item. The observed ID values can differ from zero because of accounting errors, measurement errors, and/or losses or diversions. We assume that the total accounting and measurement error for any ID measurement is normally distributed with a mean of zero and, based on historical data, a standard deviation of $\sigma = 0.4$ kilograms (kg). We wish to test whether there have been any losses or diversions. To ensure protection against a serious loss or diversion, we choose a strategic quantity of 1 kg, and the facility is required to show that any observed loss from the ID measurements is significantly smaller than 1 kg.

Suppose that $n = 12$ independent ID measurements are made in a given month and the average ID is calculated as $\bar{y} = 0.66$. We test the null hypothesis of $\mu_0 = 1$ against a left-sided alternative. The setup for this test is:

$$H_0: \mu = \mu_0 = 1, H_1: \mu < 1, \sigma = 0.4, n = 12, \alpha = 0.05.$$

The test statistic is:

$$z = \frac{\bar{y} - 1.00}{0.4/\sqrt{12}} = \frac{0.66 - 1.00}{0.4/\sqrt{12}} = -2.94$$

Since z is less than the critical value $z_{0.05} = -1.645$, we reject the null hypothesis and conclude that any loss experienced by the facility is less than the strategic quantity of 1 kg (i.e., the facility is in compliance). If the test statistic had been larger than the critical value, we would not have rejected the null hypothesis and concluded that the facility did not prove compliance.

13.4 Power of a test of hypothesis

If the null hypothesis is not rejected, the temptation is to accept it. However, this action is generally improper and ill advised, as acceptance is not necessarily the other side of rejection. As a wise man once said, an absence of evidence is not evidence of an absence. We are ready to elucidate.

First, consider $Z = (\bar{Y} - \mu_0)\sqrt{n}/\sigma$ from Equation (13.3). Using Z to test H_0 against a right-sided alternative, we reject H_0 if Z is greater than the critical value $z_{1-\alpha}$. Equivalently, there is a value \bar{y}_0 of \bar{Y} such that we reject H_0 if \bar{y} is greater than \bar{y}_0 . To find \bar{y}_0 , set Z to $z_{1-\alpha}$ and solve for \bar{y}_0 to get:

$$\bar{y}_0 = \mu_0 + z_{1-\alpha}\sigma/\sqrt{n} \quad (13.4)$$

Returning to Example 13.1, we substitute $\mu_0 = 10$, $\sigma = 3$, $n = 100$, and $z_{1-\alpha} = 1.645$ to calculate \bar{y}_0 as:

$$\bar{y}_0 = 10 + (1.645)(3)/\sqrt{100} = 10.494$$

Thus, in Example 13.1, we reject H_0 if $\bar{y} > 10.494$. We feel comfortable with this decision because we know that the critical value of 10.494 is chosen so that the probability of making a Type I error (see Section 10.3) and being wrong when H_0 is true is $\alpha = 0.05$. However, the same cannot be said about *not* rejecting H_0 (i.e., accepting H_0) if $\bar{y} \leq 10.494$, because the probability of making a Type II error and being wrong when H_0 is false is not known. From Section 10.3, this probability is denoted by β and depends on the true value of μ ,

which is unknown. To evaluate the possible consequences of a test of hypotheses, it is necessary to calculate the operating characteristic of the test, $\beta = \beta(\mu)$, or the power of the test, $1 - \beta = 1 - \beta(\mu)$, as a function of the possible values of μ .

power of the test The **power of the test** (or simply the **power**) is the probability of rejecting H_0 as a function of μ . We have:

$$\text{Power}(\mu) = 1 - \beta(\mu) = \Pr\{\bar{Y} > \bar{y}_0\} \quad (13.5)$$

In general, the sample mean \bar{Y} is normally distributed with a mean μ and standard deviation σ/\sqrt{n} . From Equation (13.4), we have:

$$\begin{aligned} \text{Power}(\mu) &= \Pr\left\{\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} > \frac{\bar{y}_0 - \mu}{\sigma/\sqrt{n}}\right\} \\ &= \Pr\left\{Z > z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right\} \end{aligned} \quad (13.6)$$

where Z is the standard normal evaluated in Table T-1 of the appendix. When $\mu = \mu_0$, Equation (13.6) becomes:

$$\text{Power}(\mu_0) = \Pr\{Z > z_{1-\alpha}\} = 1 - \Pr\{Z < z_{1-\alpha}\} = 1 - (1 - \alpha) = \alpha \quad (13.7)$$

➤ Equation (13.7) is equivalent to the definition of the level of significance.

Note that Equation (13.6) is the power for testing H_0 against a right-sided alternative. To test H_0 against a left-sided alternative, the critical value $z_{1-\alpha}$ is replaced by z_α and each greater-than symbol ($>$) is replaced by a less-than symbol ($<$) in Equations (13.5) and (13.6). For a left-sided alternative, the power is given by:

$$\text{Power}(\mu) = \Pr\left\{Z < z_\alpha + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right\} \quad (13.8)$$

To test H_0 against a two-sided alternative, there are two critical values and the power is a combination of Equations (13.6) and (13.8), using $\alpha/2$ and $1 - \alpha/2$, for the two critical values.

Continuing with Example 13.1, the probability of rejecting H_0 , which is the probability that \bar{Y} is larger than 10.494, is illustrated in Figure 13.1. In that figure, the normal density on the left is the distribution of \bar{Y} under H_0 with its center at $\mu_0 = 10$ and a critical point at 10.494. The dark shaded area to the right of the critical point represents α , which in our case is $\alpha = 0.05$.

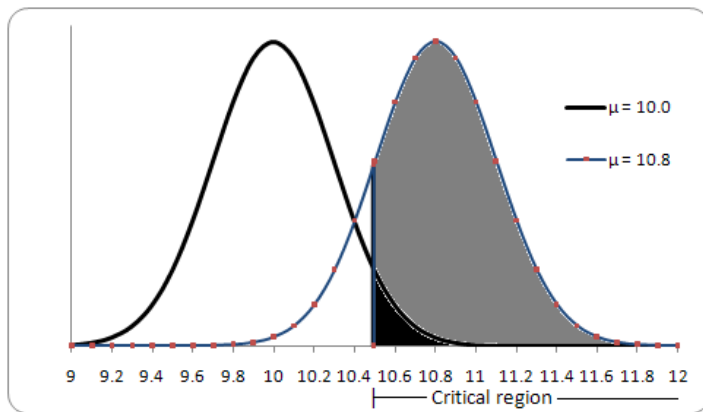


Figure 13.1. Probability of rejection of the null hypothesis

Now suppose that $\mu = 10.8$. The normal density on the right side of Figure 13.1 is the distribution of \bar{Y} for $\mu = 10.8$. The probability of rejecting H_0 is represented by the area under that curve to the right of the critical point of 10.494. This probability, schematically shown as the sum of the dark and light shaded areas, is calculated as:

$$\begin{aligned} Pr\{10.494 < \bar{Y}\} &= Pr\left\{\frac{10.494 - 10.80}{3/10} < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}\right\} \\ &= Pr\{Z > -1.020\} = Pr\{Z < 1.020\} = 0.846 \end{aligned}$$

Thus, the power of the test for $\mu = 10.8$ is $1 - \beta(10.8) = 0.846$. The probability of a Type II error is $\beta = \beta(10.8) = 0.154$.

Using Equation (13.6), we can calculate the power of the test for any other value of μ . Table 13.1 shows the power for selected values of μ .

Table 13.1. Power of the test for Example 13.1

μ	Power, $1 - \beta$	β
10.0	0.050	0.950
10.2	0.164	0.836
10.4	0.377	0.623
10.494	0.500	0.500
10.6	0.638	0.362
10.8	0.846	0.154
11.0	0.954	0.046
11.2	0.991	0.009
11.4	0.999	0.001
11.6	1.000	0.000

power curve

Figure 13.2 shows a graph of the power function, called a **power curve**, as given by Equation (13.6). Note that when $\mu = \mu_0 = 10$, the power is $\alpha = 0.05$ and that when $\mu = z_{1-\alpha} = 10.494$, the power is 0.50. A key feature of the power curve is that it increases as a function of μ and asymptotically approaches 1. This follows directly from Equation (13.6) where $Power(\mu)$ approaches 0 as μ approaches $-\infty$ and $Power(\mu)$ approaches 1 as μ approaches ∞ . This makes sense because the larger μ is than μ_0 , the larger the probability that the test will “sense” that H_0 is incorrect and reject it. For this reason, the power of the test is sometimes referred to as the sensitivity of the test.

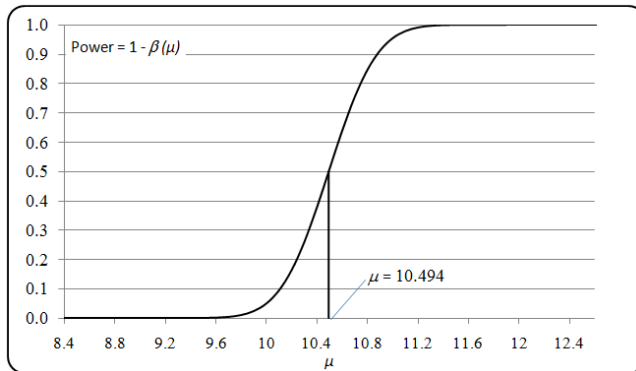


Figure 13.2. Power curve for Example 13.1

13.5 Operating characteristic curve

*operating
characteristic
(OC) curve*

The power curve is a plot of the probability of rejecting H_0 as a function of the true value of μ . In many investigations, however, it is more convenient to plot the probability of accepting H_1 (i.e., not rejecting H_0 as a function of μ). That plot is called the *operating characteristic curve*, or *OC curve*. Mathematically, the OC curve is 1 minus the power of the test. The last column of Table 13.1 shows selected values of the OC for Example 13.1. The corresponding OC is plotted in Figure 13.3.

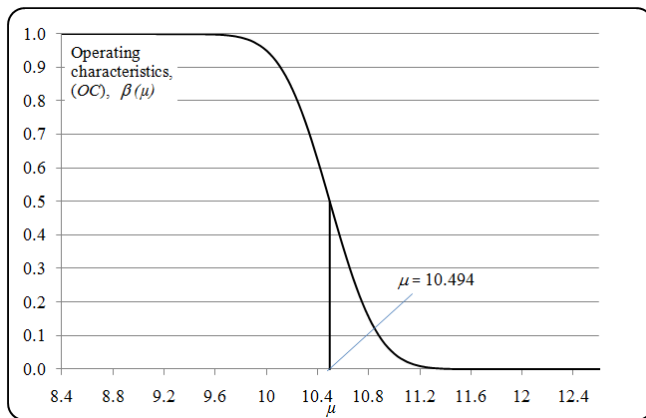


Figure 13.3. Operating characteristic curve for Example 13.1

13.6 More power to you

If we are concerned that we may fail to reject a null hypothesis that is false, we must address the power of the test and see where that power can be improved. We return to Example 13.1 and its associated power curve depicted in Figure 13.2.

The power of the test quantifies the sensitivity of the test to departures from the null hypothesis. What can we do to make the test more sensitive? Three options are considered here, although we recognize that these options are not always practical or available: (1) increase the probability of a Type I error, (2) increase the sample size, or (3) decrease the standard deviation.

Option 1. Increase the probability of a Type I error. As α increases, β necessarily decreases. As we increase α , we increase the size of the critical region, thereby increasing the probability $1 - \beta$ of the test statistic falling into the critical region.

Option 2. Increase the sample size. As the sample size n increases, from Equation (13.6), the power of the test also increases. Of course, large samples are more costly and may even be impossible to obtain at any cost. Also, the increase in power is not linearly proportional to the sample size. For example, increasing the sample size from 100 to 200 has far less effect on the power than increasing the sample size from 10 to 20.

Option 3. Decrease the standard deviation. As the standard deviation σ decreases, from Equation (13.6), the power of the test increases. In a manufacturing process, this requires tighter quality control, by controlling or accounting for more factors that may affect the product or process.

13.7 Testing a mean when σ is unknown

Rarely will we be fortunate enough to know the true standard deviation, σ , which is associated with a population that we examine. The steps necessary to test a hypothesis about a population mean, however, are natural extensions of the case where σ is known. As we might expect, given a sample of size n , σ is estimated by the sample standard deviation S with $(n - 1)$ degrees of freedom.

From Equation (13.3), the test statistic used when σ is known is Z . Our intuition tells us that the σ in the denominator of the Z statistic should be replaced by S . The test statistic thus constructed, however, is *not* normally distributed; it has a Student's t -distribution with $(n - 1)$ degrees of freedom. The test statistic is:

$$T = \frac{\bar{Y} - \mu}{S / \sqrt{n}} = \frac{\bar{Y} - \mu}{S} \sqrt{n} \quad (13.9)$$

Once the sample data are collected, the values of \bar{Y} , S , and n and the value of μ stated under the null hypothesis are substituted in Equation (13.9). The calculated value t is then compared to the appropriate quantile of the t -distribution, denoted by $t_{1-\alpha}(n - 1)$, in Table T-3 of the appendix for $df = n - 1$ and $q = 1 - \alpha$.

Example 13.3. Fuel pellets and solid lubricant additive. The percent of solid lubricants added to fuel composition before compaction into pellets is recorded for a random sample of size 7 as 1.13, 0.93, 1.06, 1.08, 1.11, 1.19, and 0.97. Does this sample support a claim that the mean percent of solid lubricant does not exceed 1.00?

This is a one-sided test, where $H_0: \mu = 1.00$ and $H_1: \mu > 1.00$. We have $\alpha = 0.05$, $n = 7$, $df = 6$, and $t_{0.95}(6) = 1.94$.

We calculate: $\bar{y} = 1.0671$, $s = 0.0907$, and

$$t = \frac{1.0671 - 1.00}{0.0907 / \sqrt{7}} = 1.956$$

Since the calculated test statistic is larger than $t_{0.95}(6)$, the null hypothesis is rejected. We conclude that the statistical evidence does not support the claim that $\mu \leq 1.00$.

Example 13.4. Ice weight in an ice condenser reactor. We return to Example 9.3 and test whether the average ice basket weight μ is at least 900 kg, so that enough ice is available for cooling in case of accident. A random sample of $n = 25$ baskets is collected. Investigate whether the reactor is in compliance.

To test for compliance, we resort to the test strategy in Section 13.3 and use an alternative hypothesis of $\mu > 900$. Also to conclude compliance, the null hypothesis would need to be rejected. The test setup is:

$H_0: \mu = 900, H_1: \mu > 900, \alpha = 0.05, n = 25, \nu = 24$, and, from Table T-3, $t_{0.95}(24) = 1.71$.

From Table 9.2, the sample mean is $\bar{y} = 926.52$, and the sample standard deviation is $s = 20.51$. Using Equation (13.6), the test statistic is

$$t = \frac{926.52 - 900}{20.51 / \sqrt{25}} = 6.47$$

Since $t = 6.47 > t_{0.95}(24) = 1.71$, H_0 is rejected and we conclude that the reactor is in compliance. Note that this conclusion is consistent with the 95% confidence interval for μ in Example 9.3, where the lower limit for μ is 919.5.

13.8 Hypotheses with two-sided alternatives

On some occasions, we will find that the alternative hypothesis is best written as a two-sided alternative. In that case, we have a critical region composed of disjoint parts. Such a critical region guards against a mean that is too large or too small, as compared with the mean specified by H_0 . A test that uses a two-sided alternative hypothesis is illustrated by the following example.

Example 13.5. Piston diameter. Twelve pistons are taken at random intervals from an assembly line, where the average piston diameter is claimed to be 8.0 centimeters (cm). The 12 observations are 7.7, 7.9, 8.4, 8.0, 8.1, 8.3, 7.9, 8.2, 8.4, 7.8, 8.0, and 8.0. The mean of these observations is $\bar{y} = 8.058$.

The hypotheses for this test are $H_0: \mu = 8.0$ and $H_1: \mu \neq 8.0$. Since the test is two-sided, we have two critical values: one at the $\alpha/2$ quantile and one at the $(1 - \alpha/2)$ quantile of the distribution of the test statistic.

We consider two cases: one when σ is assumed known and one when σ is not known.

Assume that $\sigma = 0.25$. From Equation (13.3), the test statistic is calculated as:

$$z = \frac{8.058 - 8.00}{0.25 / \sqrt{12}} = 0.804$$

Since we have a two-sided test, we look up the $1 - \alpha/2$ quantile of the standard normal distribution (Table T-1) where $z_{0.975} = 1.960$. Since $|z| < 1.960$, we do not reject H_0 and conclude that the data support the claim that the average piston diameter is 8.0 cm.

If σ is not known, we estimate σ from the sample by $s = 0.2275$. From Equation (13.6), we have:

$$t = \frac{8.058 - 8.00}{0.2275 / \sqrt{12}} = 0.883$$

Having a two-sided test, we look up the $1 - \alpha/2$ quantile of the Student's t -distribution (Table T-3) to find $t_{0.975}(11) = 2.20$. Since $|t| < 2.20$, we do not reject H_0 and conclude that the data support the claim that the average piston diameter is 8.0 cm. This is the same conclusion we reached when σ is assumed to be 0.25.

13.9 Required sample size to test a mean: σ known

The subject of sample size is first introduced in Section 9.14 where a confidence interval for the mean μ of a normal population at a specified level of $100(1 - \alpha)\%$ is required. In this case, σ is assumed known and a margin of error d is specified. The required sample size n as a function of α , σ , and d is given by Equation (9.21), which is copied here for reference.

$$n = \left(\frac{z_{\alpha/2} \sigma}{d} \right)^2 = z_{\alpha/2}^2 \left(\frac{\sigma}{d} \right)^2 \quad (13.10)$$

For hypothesis testing, the role that the margin of error plays for confidence intervals is replaced by the distance between μ_0 and a specified alternative value of μ . To see why this is so, consider the power function given by Equation (13.6) for testing $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$. What happens as the sample size n increases? Suppose first that H_1 is true so that $\mu > \mu_0$. As $n \rightarrow \infty$, $\text{Power}(\mu) \rightarrow \Pr\{Z > -\infty\} = 1$. Now suppose that $\mu < \mu_0$. Then, as $n \rightarrow \infty$, $\text{Power}(\mu) \rightarrow \Pr\{Z > \infty\} = 0$. Both of these limiting results make sense because, as the sample size increases, the sample mean \bar{Y} approaches its mean μ . Finally,

if $\mu = \mu_0$, $Power(\mu) = Pr\{Z > z_{1-\alpha}\} = \alpha$. In summary, as $n \rightarrow \infty$, the power curve approaches the ideal power curve, which is a step function equal to 0 for $\mu < \mu_0$ and 1 for $\mu > \mu_0$. Figure 13.4 illustrates this.

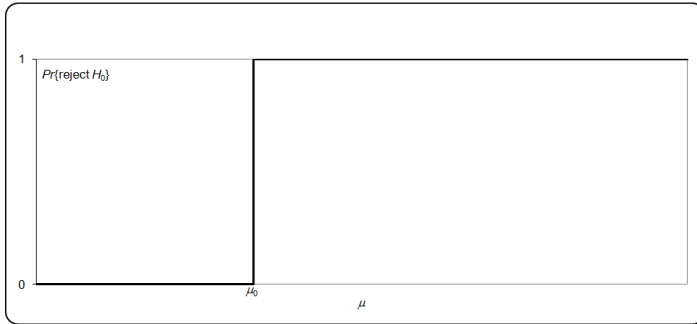


Figure 13.4. Ideal power curve for a right-sided alternative

For finite n , the power curve can at best approximate the ideal power curve, as illustrated in Figure 13.5. In practice, in addition to specifying the Type I error α , the Type II error β is often also specified for a specified alternative $\mu_1 > \mu_0$. In other words, besides being equal to α when $\mu = \mu_0$ (see Equation (13.7)), the power curve is constrained to be equal to $1 - \beta$ when $\mu = \mu_1$, as indicated in Figure 13.5.

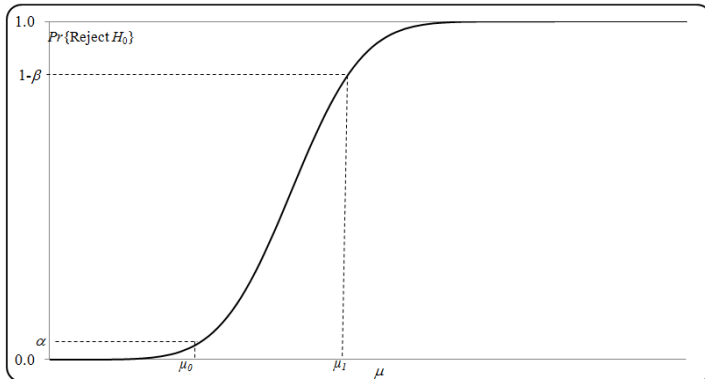


Figure 13.5. Power curve for specified Type I and Type II errors

Given μ_0 and α , the sample size n is determined once μ_1 and β are specified. From Equation (13.5):

$$Power(\mu_1) = 1 - \beta(\mu_1) = 1 - \beta = Pr\{Z < z_{1-\beta}\} = Pr\{Z > -z_{1-\beta}\} \quad (13.11)$$

where Z is the standard normal. From Equation (13.6), we have

$$Power(\mu_1) = Pr \left\{ Z > z_{1-\alpha} + \frac{(\mu_0 - \mu_1)\sqrt{n}}{\sigma} \right\} \quad (13.12)$$

next, equate the right-hand sides of Equations (13.11) and (13.12) and solve for n to get:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})\sigma^2}{(\mu_1 - \mu_0)^2} \quad (13.13)$$

- Because sample sizes are integers, a noninteger value calculated from Equation (13.13) is rounded up to the next largest integer.
- Although derived from the power function for a right-sided alternative, Equation (13.13) also holds for a left-sided alternative. Furthermore, if α is replaced by $\alpha/2$, it is a good approximation of the required sample size for a two-sided test.
- Comparison of Equations (13.10) and (13.13) reiterates that the role of the margin of error d for a confidence interval is played by $|\mu_1 - \mu_0|$, the distance between the null hypothesis and the specified alternative.

Example 13.6. Response time to alarm, continued. We revisit Example 13.1. Suppose we wish to determine n so that the power at $\mu_1 = 10.4$ is 0.95. We have $\alpha = \beta = 0.05$, $\sigma = 3$, and $|\mu_1 - \mu_0| = 0.4$. Substituting these values into Equation (13.13) yields $n = 608.9$, which is rounded up to $n = 609$. If we relax our power requirements and set $\mu_1 = 11$, then the required sample size is 97.4, rounded up to $n = 98$. Hence, the sample size of $n = 100$ in Example 13.1 is barely adequate to meet this power requirement.

If σ is not known, then, as for a confidence interval, the required sample size could be determined if the ratio $r_H = |\mu_1 - \mu_0|/\sigma$ could be specified. If $\alpha = \beta = 0.05$, from Equation (13.13), we have:

$$n = \frac{(1.645 + 1.645)^2}{r_H^2} = \frac{10.8}{r_H^2}$$

13.10 Required sample size to test a mean: σ unknown

When σ is unknown, we use a two-stage procedure to make a preliminary estimate of σ and then use this estimate to find the required sample size. This approach is completely analogous to the procedure used in Section 9.15 for confidence intervals.

As in Section 13.9, suppose we wish to test $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ for a specified α and with power equal to $1 - \beta$ when $\mu = \mu_1$ for specified β and $\mu_1 > \mu_0$. We use a two-stage test based on the t -distribution.

Stage 1: Take an initial random sample of size n_1 of at least three observations. Calculate the sample mean, denoted by \bar{Y}_1 and the sample standard deviation, denoted by s_1 . Next, use Table T-3 of the appendix to find $t_{1-\alpha}(n_1 - 1)$ and $t_{1-\beta}(n_1 - 1)$. Determine n from Equation (13.14) as:

$$n = \frac{\left[t_{1-\alpha}(n_1 - 1) + t_{1-\beta}(n_1 - 1) \right]^2 s_1^2}{(\mu_1 - \mu_0)^2} \quad (13.14)$$

Because sample sizes must be integers, n is rounded up to the next largest integer.

Stage 2: Let $n_2 = n - n_1$. If $n_2 \leq 0$, no additional observations are required. Otherwise, take n_2 additional observations and denote the mean of this second sample by \bar{Y}_2 .

Let \bar{Y} be the sample mean of the $n = n_1 + n_2$ observations. Then \bar{Y} may be calculated as:

$$\bar{Y} = \frac{n_1 \bar{Y}_1 + n_2 \bar{Y}_2}{n_1 + n_2} \quad (13.15)$$

The critical region for testing $H_0: \mu = \mu_0$ against a right-sided alternative is:

$$\bar{Y} > \mu_0 + t_{1-\alpha}(n_1 - 1) \frac{s_1}{\sqrt{n}} \quad (13.16)$$

With this two-stage procedure, the power when $\mu = \mu_1$ is at least $1 - \beta$.

14

Tests for statistical hypotheses: Variances

14.1 What to look for in Chapter 14

Chapter 14 provides a methodology for testing various hypotheses about variances. The discussion begins with an exploration of testing hypotheses about a single variance, then moves on to compare two variances, and concludes with a procedure for comparing several variances. Topics to look for are:

- *chi-square statistic for testing a single variance, §14.3*
- *one- and two- sided tests of a single variance, §14.4*
- *F statistic for testing equality of two variances, §14.5*
- *F_{max} statistic for testing homogeneity of several variances, §14.5*
- *pooling variances, §14.6*
- *Bartlett's test for homogeneity of several variances, §14.7*

14.2 Why worry about variances?

Any production line that claims output quality must maintain control over the variability of its product, as well as its average value. It is simply not enough to have good average quality; it is equally important for quality variability to be controlled. Recall the context of the precision of a measurement system from Chapter 6. The variance is a measure of precision. A precise measurement system is one with a small variance.

- If a manufacturer knows both the variability as well as the average life expectancy of his product, he can use this information to select an appropriate period to warrant his product.

From a statistical perspective, many statistical procedures make specific assumptions about the variance of a population or about the relationships between the variances of several populations. If these assumptions are inappropriate, so are the procedures. These assumptions are often checked by using techniques described in this chapter. Thus, being able to assume that the variances are equal allows us to use analytical techniques such as the analysis of variance (Chapter 16). If the variances cannot be assumed equal, we need to rethink the analysis and, perhaps, use a nonparametric statistical procedure (Chapter 25).

This chapter covers three distinct scenarios in which variances are subjected to statistical tests. These tests are used to determine whether:

- A population variance equals a given value.
- Two populations have equal variances.
- Several populations have equal variances.

Throughout this chapter, all populations are assumed to be normally distributed with unknown means.

- Note that every test presented in this chapter addresses a question about a population variance (or variances) rather than a population standard deviation. However, because of their functional relationship, any inference about a variance is automatically applicable to the standard deviation.
- Because no assumptions or claims are made about the means of the populations whose variances we test, we must estimate these means because our variance tests are all based on the sample variances. Otherwise, the sample means are not explicitly used in the variance tests. Contrast this with tests of hypotheses about a mean (Chapter 13) or means (Chapters 16 and 17), where assumptions or estimates about the variances are an integral part of the tests.

14.3 Testing a single variance

We have a sample of size n . The null hypothesis about a specified value σ_0^2 of the variance is written as:

$$H_0: \sigma^2 = \sigma_0^2 \quad (14.1)$$

The alternative hypothesis can take one of the following three forms:

$$H_1: \sigma^2 > \sigma_0^2 \quad (14.2)$$

$$H_1: \sigma^2 < \sigma_0^2 \quad (14.2a)$$

$$H_1: \sigma^2 \neq \sigma_0^2 \quad (14.2b)$$

The choice among the three alternative hypotheses is not always obvious. Considerations in making this choice depend on our priorities and concerns. However, Equation (14.2) is usually the alternative hypothesis because we are generally concerned only about variances that are too large (i.e., larger than the hypothesized value).

The formulation of the problem is completed by stating the desired level of significance. While the choice of the level of significance also depends on priorities and concerns, this book generally uses the default value of $\alpha = 0.05$.

In testing whether $\sigma^2 = \sigma_0^2$, we compare the sample variance S^2 to the hypothesized σ^2 . Equation (14.3) gives the test statistic.

$$\chi^2(n-1) = \frac{(n-1)S^2}{\sigma_0^2} \quad (14.3)$$

Under the null hypothesis, the test statistic in Equation (14.3) has a chi-square distribution with $\nu = n - 1$ degrees of freedom, the same number of degrees of freedom that is associated with S^2 . The critical point $\chi_{1-\alpha}^2(\nu)$ for the alternative hypothesis in Equation (14.2) is obtained from Table T-2 of the appendix, although some interpolation of the table may be needed. Alternatively, the critical point may be obtained using Excel's function =CHIINV(1 - α , ν). If the calculated test statistic $\chi^2(\nu)$ exceeds $\chi_{1-\alpha}^2(\nu)$, H_0 is rejected, and we conclude that the population variance is larger than the hypothesized σ_0^2 .

Example 14.1 illustrates this test.

Example 14.1. Testing a single variance. A manufacturer claims that the standard deviation associated with construction steel rod length is no more

than $\sigma = 10$ millimeters (mm) (or, equivalently, $\sigma^2 = 100 \text{ mm}^2$), a claim we need to verify. A random sample of $n = 20$ observations yields a sample variance of 177.13 mm. The hypotheses for this test are written as $H_0: \sigma^2 = 100$ and $H_1: \sigma^2 > 100$.

The calculated test statistic is:

$$\chi^2(19) = \frac{(19)(177.13)}{100} = 33.65$$

From Table T-2, the critical point for the test is $\chi_{0.95}^2(19) = 30.1$. Excel yields $=\text{CHIINV}(0.05, 19) = 30.14$.

Since the calculated chi-square exceeds the critical point, we reject H_0 and conclude that the statistical evidence does not support the claim that $\sigma^2 \leq 100$.

Had the alternative hypothesis been $\sigma^2 < 100$, the critical point would have been $\chi_{0.05}^2(19) = 10.1$ and the test would have rejected H_0 if the calculated $\chi^2(19)$ were less than 10.1.

Had the alternative hypothesis been $\sigma^2 \neq 100$, then two critical points would have been needed. The lower critical point would have been $\chi_{0.025}^2(19) = 8.91$ and the higher critical point would have been $\chi_{0.975}^2(19) = 32.9$. If the calculated test statistic had been larger than 32.9 or smaller than 8.91, then H_0 would have been rejected.

An alternative to testing $H_0: \sigma^2 = \sigma_0^2$ appears in Chapter 9, where a confidence interval for σ^2 is constructed. If the confidence interval does not include σ_0^2 , H_0 must be rejected. We illustrate this by constructing a two-sided confidence interval for σ^2 in Example 14.1. From Equation (9.18), we have the following confidence interval for σ^2 :

$$\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right) = \left(\frac{(20-1)177.13}{32.9}, \frac{(20-1)177.13}{8.91} \right) = 102.3, 377.7$$

Since the confidence interval does not include $\sigma_0^2 = 100$, H_0 is rejected.

14.4 Testing equality of two variances

Many statistical procedures are based on the assumption that the variances of two or more populations are equal. In this section, we show how to test such an

assumption for two populations. Sections 14.5 and 14.7 show how to test the equality of variances when more than two populations are involved.

Let σ_A^2 and σ_B^2 denote the variances of a product manufactured by Manufacturers A and B, respectively. Although neither variance is known, we wish to test whether they are equal. We start with the null hypothesis:

$$H_0: \sigma_A^2 = \sigma_B^2 \quad (14.4)$$

The alternative hypothesis can take one of the following three forms:

$$H_1: \sigma_A^2 > \sigma_B^2 \quad (14.5)$$

$$H_1: \sigma_A^2 < \sigma_B^2 \quad (14.5a)$$

$$H_1: \sigma_A^2 \neq \sigma_B^2 \quad (14.5b)$$

Depending on the context, any one of these alternatives may be appropriate. If we just want to test the equality of the two variances, we use a two-sided alternative hypothesis, namely, $H_1: \sigma_A^2 \neq \sigma_B^2$.

As usual, we set $\alpha = 0.05$.

From the first population, we draw a random sample of n_A observations and calculate the sample variance S_A^2 with $\nu_A = n_A - 1$ degrees of freedom. From the second population, we draw a random sample of n_B observations and calculate the sample variance S_B^2 with $\nu_B = n_B - 1$ degrees of freedom.

Let S_{max}^2 and S_{min}^2 denote the larger and smaller of S_A^2 and S_B^2 respectively.

Likewise, let ν_{max} and ν_{min} denote the degrees of freedom of S_{max}^2 and S_{min}^2 , respectively.

F statistic The test statistic for examining equality of two variances is the
F ratio ***F statistic*** (also called the ***F ratio*** or ***variance ratio***). That
variance statistic is:
ratio

$$F = S_{max}^2 / S_{min}^2 \quad (14.6)$$

Even though the alternative is two-sided, we have only one critical value because, as constructed, the F statistic cannot be smaller than 1. We use Table T-4 of the appendix to obtain the critical value and reject H_0 if the critical value is exceeded. The table requires three input variables: quantile, ν_{max} , and ν_{min} . Since the alternative hypothesis is two-sided, the critical value is the

$(1 - \alpha/2)^{\text{th}} = 97.5^{\text{th}}$ quantile of the *F-distribution*, denoted by $f_{0.975}(v_{\max}, v_{\min})$, where the degrees of freedom in the numerator is v_{\max} , and the degrees of freedom in the denominator is v_{\min} .

- Table T-4 is rather sparse and interpolation is often necessary.
- Alternatively, we can use Excel's function =FINV($\alpha/2$, v_{\max} , v_{\min}) to find the critical value.

Example 14.2. Testing equality of two variances. Suppose we are given summary statistics calculated from two samples, A and B, as shown in Table 14.1.

Table 14.1. Testing equality of variance for two populations

Sample	Mean	Variance	n	$\nu = n - 1$
A	18.5	207	16	15
B	21.6	317	25	24

From Table 14.1, $S_{\max}^2 = 317$, $S_{\min}^2 = 207$, $v_{\max} = 24$, and $v_{\min} = 15$. The variance ratio is $f = 317/207 = 1.531$. From Table T-4, we find $f_{0.975}(24, 15) = 2.70$. Because $1.531 < 2.70$, we do not have sufficient evidence to conclude that the variances are different.

- Note that the sample means are not explicitly used in this test.

The following example shows how to test a one-sided alternative.

Example 14.3. Warehouse temperature variation. Temperature variation in a local warehouse is measured by its standard deviation. Management will consider switching to a remote warehouse if it can be shown that its standard deviation is lower than that of the local warehouse. A sample of $n_L = 50$ ($\nu_L = 49$) temperature observations (in F°) at the local warehouse yields a sample variance of $S_L^2 = 43.56$ ($S_L = 6.6$), and a sample of $n_R = 60$ ($\nu_R = 59$) temperature observations at the remote warehouse yields a sample variance of $S_R^2 = 11.56$ ($S_R = 3.4$). The null and alternative hypotheses for this test are:

$$H_0: \sigma_L^2 = \sigma_R^2, \quad H_1: \sigma_L^2 > \sigma_R^2, \quad \alpha = 0.05$$

The test statistic for this example is $f = 43.56/11.56 = 3.768$. Because this is a one-sided test, the quantile used here is $(1 - \alpha) = 0.95$. Based on interpolation in Table T-4, the critical value is $f_{0.95}(49, 59) = 1.57$. The same value may also be obtained from Excel's function =FINV(0.05, 49, 59). Since f exceeds the critical value, we reject H_0 and conclude that the remote warehouse has less temperature variation than the local warehouse.

14.5 Testing homoscedasticity with samples of equal size

balanced experiment In Section 14.4, we learned how to test whether the variances of two populations are equal. The remainder of this chapter is devoted to tests of equality of variances when there are two or more populations. This section provides a test for a **balanced experiment** (i.e., an experiment where all sample sizes are the same). Section 14.7 shows how to test with an unbalanced experiment.

homoscedasticity A term used to describe equality of variances is **homoscedasticity**, and a term used to describe inequality of variances is **heteroscedasticity**.

robust Many statistical procedures that assume equality of variances are not very strict about this assumption because these procedures are **robust** (i.e., the test results are not appreciably affected by minor departures from homoscedasticity). Accordingly, we will sometimes assume that the variances are similar or alike (i.e., they are approximately equal). For the same reason, tests for homoscedasticity are often conducted at the 0.01 level of significance, because such tests are less sensitive to a departure from homoscedasticity. In this Chapter, however, we will continue to use the default level of $\alpha = 0.05$.

F_{max} Tests of homoscedasticity are typically lengthy and computationally tedious. However, when the experiment is balanced, the following procedure, called the F_{max} test, may be used.

Hartley's test The F_{max} test, also known as **Hartley's test**, tests the homoscedasticity of k populations, $k \geq 2$, where the sample sizes are equal (i.e., $n_1 = n_2 = \dots = n_k = n$). A reference to this test may be found in Beyer (1974), pp. 328–329. The null hypothesis is:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \quad (14.7)$$

The alternative hypothesis is:

$$H_1: \sigma_i^2 \neq \sigma_j^2, \quad i \neq j \text{ for at least one pair } (i, j), \quad i \neq j \quad (14.8)$$

The level of significance for this test is set at $\alpha = 0.05$.

The test consists of the following steps:

Step 1. Calculate the sample variances for each of the k samples, where each sample variance has $\nu = n - 1$ associated degrees of freedom.

Step 2. Set $S_{max}^2 =$ the largest of the k sample variances.

Step 3. Set $S_{min}^2 =$ the smallest of the k sample variances.

Step 4. Calculate F_{max} by:

$$F_{max} = S_{max}^2 / S_{min}^2 \quad (14.9)$$

Step 5. Find $f_{max,q}(k, \nu)$, the $q^{th} = (1 - \alpha)^{th}$ quantile, of the F_{max} distribution for k groups and ν degrees of freedom in in Table T-5, “Quantiles, $f_{max,q}(k, \nu)$, for the F_{max} distribution,” of the appendix.

Step 6. Reject H_0 if f_{max} , the calculated value of F_{max} , exceeds $f_{max,q}(k, \nu)$, and conclude that there is statistical evidence of heteroscedasticity. Otherwise, we do not have sufficient evidence to conclude that the variances are not all the same.

Strictly speaking, each of the k samples must have exactly n observations. However, for most applications, this test is quite adequate if the number of observations is roughly equal. The degrees of freedom for the test is set equal to the average number of degrees of freedom for the k samples.

Example 14.4. Testing equality of uranium concentration

variances. A study was conducted to see if materials supplied by six different uranium dioxide (UO₂) suppliers were homoscedastic with respect to their uranium concentration. Random samples of size $n = 8$ each were collected from each of the UO₂ suppliers. Table 14.2 summarizes the data.

Table 14.2. Uranium concentration (percent) in UO₂ samples

Supplier	Mean, \bar{y}	Variance, S^2	Size, n	$\nu = n - 1$
1	86.2	0.40	8	7
2	88.1	1.49	8	7
3	87.3	3.20	8	7
4	89.9	1.93	8	7
5	86.2	2.35	8	7
6	89.2	1.58	8	7

From this table, we calculate $f_{max} = 3.20/0.40 = 8.00$. From appendix Table T-5, the critical value for the test is $f_{max,0.95}(6, 7) = 10.8$. Since f_{max} is smaller than the critical value, we do not reject the null hypothesis of homoscedasticity. The

statistical evidence is insufficient to conclude that the population variances are different.

14.6 Pooling variances

pooling variances Before introducing a test for equality of variances with an unbalanced experiment, we first introduce the concept of **pooling variances**.

Suppose we are given k populations, $k \geq 2$, and have a reason, statistical or otherwise, to believe that the population variances are the same. We collect k independent samples, one from each population. Let S_i^2 , n_i , and $\nu_i = (n_i - 1)$ denote the sample variance, sample size, and degrees of freedom of the i^{th} sample, respectively. If all the population variances are indeed the same, we have k independent sample variances, each estimating the same variance. We wish to combine, or pool, all the sample variances to obtain a single estimate of the common population variance.

weighted average Pooling variances means averaging the variances. To be more exact, we form a **weighted average** of the k sample variances, where the weights are the degrees of freedom. The pooled variance is denoted by S_p^2 , or simply as S^2 , and is calculated as:

$$S^2 = \frac{\sum_{i=1}^k (n_i - 1) S_i^2}{\sum_{i=1}^k (n_i - 1)} = \frac{\sum_{i=1}^k (n_i - 1) S_i^2}{\sum_{i=1}^k n_i - k} \quad (14.10)$$

The pooled variance has ν associated degrees of freedom, where

$$\nu = \sum_{i=1}^k (n_i - 1) = \sum_{i=1}^k n_i - k \quad (14.11)$$

It can be shown that S^2 is the “best” estimator of the common population variance (i.e., it has the smallest variance of all weighted averages of the k sample variances).

➤ Note that we pool the variances, not the standard deviations.

The following example illustrates the pooling of $k = 4$ sample variances.

Example 14.5. Pooling variances. Table 14.3 summarizes the percentage of uranium contents for four samples. We calculate the pooled variance.

Table 14.3. Pooled variance for uranium content

Sample, i	1	2	3	4	Sum
Variance, s_i^2	0.026	0.054	0.063	0.031	
Size, n_i	11	15	14	16	56
$v_i = (n_i - 1)$	10	14	13	15	52
$(n_i - 1) s_i^2$	0.260	0.756	0.819	0.465	2.300

The pooled variance is:

$$S^2 = \frac{10(0.026) + 14(0.054) + 13(0.063) + 15(0.031)}{10 + 14 + 13 + 15} = 0.0442$$

14.7 Testing equality of variances with unequal sample sizes

Bartlett's test *Bartlett's test* (Bartlett, 1937) is used for testing the homoscedasticity of several populations. In contrast to the F_{\max} test, Bartlett's test does not require that all samples be of equal size. Bartlett's test is the most commonly used test for homoscedasticity but has a shortcoming in that, while sensitive to heteroscedasticity, it is also somewhat sensitive to departures from normality. This shortcoming is usually tolerated.

Levene's test A test of homoscedasticity that is not sensitive to departures from normality is *Levene's test* (Levene, 1960). This test is available in most statistical software packages.

The null and alternative hypotheses for comparing k populations, $k \geq 2$, are the same as stated in Equations (14.7) and (14.8), respectively.

Let n_i and S_i^2 denote the sample size and the sample variance for sample i , respectively, and let $N = \sum_{i=1}^k n_i$ denote the total number of observations.

Bartlett's test statistic, denoted by B , is given by:

$$B = \frac{(N - k) \ln(S^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \left(\frac{1}{n_i - 1} \right) - \frac{1}{N - k} \right)} \quad (14.12)$$

Here, S^2 is the pooled variance, defined in Equation (14.10), and \ln is the natural logarithm function.

Under H_0 , Bartlett's test statistic is distributed approximately as a chi-square variable $\chi^2(k-1)$ with $k-1$ degrees of freedom. From Table T-2 of the appendix, the critical value is $\chi^2_{1-\alpha}(k-1)$ and H_0 is rejected if it is exceeded.

Example 14.6. Continuation of Example 14.5. Table 14.3 presents the data for this example. Table 14.4, which mirrors an Excel spreadsheet, shows the calculations for Bartlett's test statistic. The entries of Table 14.3 are copied onto the first four rows of Table 14.4, (cells B2 ... E5) and we continue with the following steps:

- Step 1.** In cell B6, enter =1/B4, the reciprocal of the degrees of freedom for Sample 1. In our example, that entry is $1/10 = 0.100$. Copy that entry to cells C6..E6.
- Step 2.** In cell F6, enter =SUM(B6..E6), the sum of the reciprocals of the degrees of freedom. In our example, that entry is 0.315.
- Step 3.** In cell B7, enter =LN(B2), the logarithm of the variance of Sample 1. In our example, that entry is $\ln(0.026) = -3.650$. Copy that entry to cells C7..E7.
- Step 4.** In cell F7, enter =SUM(B7..E7), the sum of the logarithms of the sample variances. In our example, that entry is -12.807.
- Step 5.** In cell B8, enter =B4*B7, the product of the degrees of freedom and the logarithm of the sample variance. In our example, that entry is -36.497. Copy that entry to cells C8..E8.
- Step 6.** In cell F8, enter =SUM(B8..E8). In our example, that entry is -165.406.
- Step 7.** In cell C9, enter the number of samples, k . In our example, $k = 4$.
- Step 8.** In cell C10, enter =F5/F4 the pooled variance. It is the same value that was calculated in Example 14.5. Here, this entry is 0.044231.
- Step 9.** In cell C11, enter =(F4)*LN(C10), the first term in the numerator of Equation (14.11). Here, we have (52) $\ln(0.044231) = -162.153$.
- Step 10.** In cell C12, enter =C11-F8, the numerator of Equation (14.11). Here, we have $-162.153 - (-165.406) = 3.253$.

Step 11. In cell C13, enter =F6-1/(F4), the term in the large parentheses in the denominator of Equation (14.11). Here, we have $0.315 - 1/(52) = 0.296$.

Step 12. In cell C14, enter =1+1/(3*(C9-1))*C13, the denominator of Equation (14.12). Here, we have $1 + 1/((3)(4 - 1))(0.296) = 1.033$.

Step 13. In cell C15, enter =C12/C14, the calculated Bartlett's test statistic. Here, we have $B = 3.253/1.033 = 3.149$.

Table 14.4. Spreadsheet calculations for Bartlett's test

	A	B	C	D	E	F
1	Sample, i	1	2	3	4	Sum
2	Variance, s_i^2	0.026	0.054	0.063	0.031	
3	Size, n_i	11	15	14	16	56
4	$(n_i - 1)$	10	14	13	15	52
5	$(n_i - 1) s_i^2$	0.260	0.756	0.819	0.465	2.300
6	$1/(n_i - 1)$	0.100	0.071	0.077	0.067	0.315
7	$\ln(s_i^2)$	-3.650	-2.919	-2.765	-3.474	-12.807
8	$(n_i - 1) \ln(s_i^2)$	-36.497	-40.863	-35.940	-52.107	-165.406
9	k	4				
10	s^2 (pooled)	0.044231				
11	$(N - k) \ln(s^2)$	-162.153				
12	$(N - k) \ln(s^2) - \sum (n_i - 1) \ln(s_i^2)$	3.253				
13	$\sum \left(\frac{1}{n_i - 1} \right) - \frac{1}{N - k}$	0.296				
14	$1 + \frac{1}{3(k - 1)} \left(\sum \left(\frac{1}{n_i - 1} \right) - \frac{1}{N - k} \right)$	1.033				
15	B	3.149				

The critical value is $\chi_{0.95}^2(3) = 7.81$. Since $B = 3.149$ does not exceed 7.81, we do not reject H_0 , and we conclude that there is no statistical evidence of heteroscedasticity.

In this example, we might have used the F_{max} test because this test is robust if the sample sizes are approximately equal. The degrees of freedom is the average over the samples, or $52/4 = 13$. From Equation (14.9), we have:

$$f_{max} = 0.063/0.026 = 2.42$$

Interpolating in Table T-5 of the appendix, we find that the critical value $f_{max,0.95}(4,13)$ is approximately 4.5. Hence, the F_{max} test would also conclude that there is no statistical evidence of heteroscedasticity.

15

Tests of statistical hypotheses: Two means

15.1 What to look for in Chapter 15

Chapter 15 introduces tools that allow us to compare the means of two populations by examining random samples from the populations. In this chapter, we assume that the samples are either from normal distributions or that the sample sizes are large enough to allow us to invoke the central limit theorem and claim that the sample means are approximately normally distributed. We use two important concepts:

- *variance and the standard deviation of mean difference, §15.4*
- *pooled variance, §15.5*
- *sample size requirements for comparing two means, §15.8*

Procedures for comparing two means are given for the following four cases:

- *observations are paired, §15.3*
- *variances are known, §15.4*
- *variances are unknown but assumed equal, §15.5*
- *variances are unknown and unequal, §15.6*

15.2 Hypotheses about the means of two populations

We often need to compare the means of two populations. Given random samples from each population, the tests in this chapter are based on the two sample means. Furthermore, the statistical calculations of the tests are all based on the assumed normality of the sample means. Accordingly, either (1) the populations are assumed to be normally distributed, and thus the sample means are also normally distributed, or (2) the sample sizes are large enough for the Central Limit Theorem (Chapter 7) to imply that the sample means are approximately normal. If the second condition holds, then the statistical calculations of the tests are approximately correct. If neither of these conditions is met, then a nonparametric procedure (Chapter 25) may apply.

When comparing the means of two populations, we do not always test whether they are equal; we may also test whether they satisfy some other relationship. The tests of hypotheses for these comparisons may take several forms. Three commonly encountered forms are defined below.

Form 1. The means are equal and the null hypothesis is:

$$H_0: \mu_2 = \mu_1, \text{ or } H_0: \mu_2 - \mu_1 = 0 \quad (15.1)$$

The alternative hypothesis is one of the following:

$$H_1: \mu_2 - \mu_1 \neq 0 \text{ (two-sided alternative)} \quad (15.1a)$$

$$H_1: \mu_2 - \mu_1 > 0 \text{ (right-sided alternative)} \quad (15.1b)$$

$$H_1: \mu_2 - \mu_1 < 0 \text{ (left-sided alternative)} \quad (15.1c)$$

As an example, suppose we test whether the average cost of electricity changed between 2009 and 2008 and write $H_0: \mu_{09} - \mu_{08} = 0$. As the test is stated, the alternative hypothesis is written as $H_1: \mu_{09} - \mu_{08} \neq 0$. However, if we are really interested in whether the average cost of electricity increased (rather than just changed) from 2008 to 2009, then the alternative hypothesis would be $H_1: \mu_{09} - \mu_{08} > 0$.

Form 2. The two means differ by a constant c , and the null hypothesis is:

$$H_0: \mu_2 = \mu_1 + c, \text{ or } H_0: \mu_2 - \mu_1 = c \quad (15.2)$$

The alternative hypothesis is one of the following:

$$H_1: \mu_2 - \mu_1 \neq c \text{ (two-sided alternative)} \quad (15.2a)$$

$$H_1: \mu_2 - \mu_1 > c \text{ (right-sided alternative)} \quad (15.2b)$$

$$H_1: \mu_2 - \mu_1 < c \text{ (left-sided alternative)} \quad (15.2c)$$

➤ If $c = 0$, the Form 2 hypotheses are equivalent to the Form 1 hypotheses.

As an example, suppose we test the claim that the response time for the new intrusion detection system is at least 2 minutes shorter than for the old system. The null and alternative hypotheses are:

$$H_0: \mu_{\text{old}} - \mu_{\text{new}} = 2, \text{ and } H_1: \mu_{\text{old}} - \mu_{\text{new}} > 2.$$

Form 3. The second mean is a constant multiple k of the first mean. The null hypothesis is:

$$H_0: \mu_2/\mu_1 = k, \text{ or } H_0: \mu_2 - k\mu_1 = 0. \quad (15.3)$$

The alternative hypothesis is one of the following:

$$H_1: \mu_2 - k\mu_1 \neq 0 \text{ (two-sided alternative)} \quad (15.3a)$$

$$H_1: \mu_2 - k\mu_1 > 0 \text{ (right-sided alternative)} \quad (15.3b)$$

$$H_1: \mu_2 - k\mu_1 < 0 \text{ (left-sided alternative)} \quad (15.3c)$$

➤ If $k = 1$, the Form 3 hypotheses are equivalent to the Form 1 hypotheses.

As an example, suppose we test whether the average cost to move an employee from Region X to Headquarters is at least 30% more than the cost of moving from Headquarters to Region X. The null and the alternative hypotheses are written as:

$$H_0: \mu_{\text{XH}} / \mu_{\text{HX}} = 1.3, \text{ or } \mu_{\text{XH}} - 1.3\mu_{\text{HX}} = 0$$

$$H_1: \mu_{\text{XH}} / \mu_{\text{HX}} > 1.3, \text{ or } \mu_{\text{XH}} - 1.3\mu_{\text{HX}} > 0$$

This chapter discusses four procedures for comparing two means. The data structure and the assumptions made drive these procedures.

15.3 Procedure 1: Paired observations

paired observations *Paired observations* are two observations made on a member of a population after it has been subjected to two different *treatments*.
treatments Because all other causes are presumably controlled, we tend to attribute any difference in the paired observations to the effects of

the treatment. This is analogous to identical twins who are expected to grow and behave alike, unless they are treated differently or are exposed to different environments. If one of the twins receives—or is deprived of—a specific food and grows faster than his/her sibling, the difference in diet is usually identified as the cause because the growth difference is unlikely to have been to the result of a genetic difference. In a statistical context, diet is a treatment. Of course, other factors may affect the observed difference between the paired observations, just as there are environmental differences that might affect the twins' growth, but for our purposes, those other factors are considered irrelevant, unlikely, or uncontrollable. The combination of all the other factors is termed **error**. Paired observations that are expected to be alike if all individuals in a group receive the same treatment are called **blocked observations**. The number of observations in a block must be two or more.

Examples of paired observations on individuals receiving different treatments include:

- Participants who are given an injection of a drug in one arm and of another drug in the second arm
- Welds that are inspected for quality by two inspectors
- Aliquots of water that are tested for impurities by two different laboratories
- Estimates for automobile repair from two different garages

In paired observation experiments, the number of paired observations is denoted by n . An individual observation is denoted by Y_{ij} , where the first subscript, i , identifies the treatment and the second subscript, j , identifies the individual. Typically, both i and j are positive integers, with the treatments denoted by 1 and 2 and the individuals numbered sequentially from 1 to n . We use $D_j = Y_{2j} - Y_{1j}$ to denote the difference between the two observations on individual j . The observed differences $\{D_j\}$ have a distribution, and we use μ_D to denote its mean.

- The subscript that identifies a treatment need not be numeric. For example, the treatment may be denoted as A and B.

Our strategy is to calculate the sample mean, \bar{D} , and the standard deviation, S_D , of the n observed paired differences. We assume that the paired differences D_j are normally distributed.

- If Y_{1j} and Y_{2j} are normally distributed, then D_j will be normal. However, it is possible for D_j to be normal even if Y_{1j} and Y_{2j} are not normal.

The model for paired observations is written as:

$$D_j = \mu_D + \varepsilon_j \quad (15.4)$$

where $\mu_D = \mu_2 - \mu_1$. The error ε_j is assumed to be distributed normally with mean 0 and unknown variance.

The test statistic for any form of the null hypothesis is:

$$T_{n-1} = \frac{(\bar{D} - \mu_D)}{S_D / \sqrt{n}} \quad (15.5)$$

where T_{n-1} has a Student's t -distribution with $\nu = n - 1$ degrees of freedom.

As an example, suppose the null hypothesis in Section 15.2 is Form 2 with a two-sided alternative. Then the null and alternative hypotheses are:

$$H_0 : \mu_D = c, \quad H_1 : \mu_D \neq c \quad (15.6)$$

The critical value is $t_{1-\alpha/2}(n-1)$ from Table T-3 of the appendix. We reject H_0 if the absolute value of the test statistic is larger than the critical value and conclude that the statistical evidence implies that the average population difference is not equal to c .

Example 15.1. Paired scale measurements. Suppose we wish to verify the equivalence of two scales used to weigh containers of special material, by weighing each of $n = 10$ containers once on each scale. The hypotheses are given by Equation (15.6) with $c = 0$. The associated degrees of freedom for the test is $\nu = n - 1 = 9$.

Table 15.1 displays the data for the 10 pairs of measurements, in kilograms (kg), along with the relevant statistics.

The test statistic is calculated as:

$$t = \frac{\bar{d} - 0}{s_{\bar{d}} / \sqrt{n}} = \frac{-0.06 - 0}{0.067} = -0.896$$

Since $|t| < t_{0.975}(9) = 2.26$, H_0 is not rejected, and we have no reason to claim that the scales are different.

For Excel users: Click on Data, Data Analysis (in some versions of Excel, click on Tools, Data Analysis), and scroll down to “t-test Paired Two-Sample for Means.” Once a new panel is displayed, select and enter the range of each variable separately. If the first row contains the group labels, check the “Label” box, otherwise be sure that this box is not checked. Enter the level of significance (default: $\alpha = 0.05$) and the cell where we wish the output to begin. Under “Hypothesized mean difference,” enter the value of c , the hypothesized μ_D in Equation (15.6), or select 0 as the default value. Table 15. 2 shows Excel’s printout (in a slightly modified style) for Example 15.1. The relevant output for this procedure gives the calculated t -statistic, as well as the critical points for both one- and two-sided tests.

Table 15.1. Container weights, in kg, measured by two scales

Container	Scale 1	Scale 2	Difference
1	25.6	25.4	-0.2
2	21.3	21.1	-0.2
3	21.3	21.7	0.4
4	28.4	28.4	0.0
5	29.9	30.0	0.1
6	30.0	29.9	-0.1
7	23.4	23.1	-0.3
8	29.5	29.6	0.1
9	27.7	27.5	-0.2
10	21.3	21.1	-0.2
Descriptive statistics for paired differences			$n = 10$
			$\nu = 9$
			$\bar{d} = -0.06$
			$s_d = 0.21$
			$s_d^* / \sqrt{n} = 0.067$

Table 15.2. Excel output for Example 15.1

Scale	Scale 1	Scale 2
Mean	25.84	25.78
Variance	13.894	14.108
Observations	10	10
Hypothesized mean difference	0	
Degrees of freedom (ν)	9	
t -statistic	0.896	
$Pr\{T \leq t\}$, one-tail	0.197	
t critical, one-tail	1.833	
$Pr\{T \leq t\}$, two-tail	0.394	
t critical, two-tail	2.262	

The spreadsheet output provides some elementary statistics for each sample: mean, variance, sample size, and the degrees of freedom. The result t -statistic and the degrees of freedom are the same as those in Table 15.1 and so is the critical point for a two-sided alternative.

Excel also provides the critical point $t_{0.95}(9)$ for a one-sided alternative, as well as the probabilities of exceeding the calculated t values for one- and two-sided alternatives.

A common application of the t -test for paired differences is to evaluate shipper-receiver differences. Differences between the quantity shipped and the quantity received of shipments may be attributed to misstatements by the shipper, misstatements by the receiver, material lost in transit, or random error. The statistical test does not identify the source of the differences, but it may indicate when random error is an unlikely source.

The value of collecting the data in pairs, when appropriate, cannot be overemphasized. Two advantages are listed below:

- The analysis of paired data is more sensitive to the population difference than an analysis that considers the data as coming from two independent samples.
- The calculations used to analyze paired observations are considerably simpler than those used for two independent samples. Indeed, pairing the observations allows us to analyze the data as if we have a single sample (of differences), a procedure covered in Chapter 13.

15.4 Procedure 2: Variances known

This procedure tests whether two normal populations, denoted here by A and B , have the same mean when the variances of the two populations, σ_A^2 and σ_B^2 , are known. The test statistic is constructed using the standard deviation of a difference in means derived as Equation (6.49) and presented below:

$$\sigma_{\bar{Y}_A - \bar{Y}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \quad (15.7)$$

where \bar{Y}_A and \bar{Y}_B are the sample means and n_A and n_B are the sample sizes for populations A and B , respectively.

From Equation (15.2), the null and alternative hypotheses for comparing μ_A and μ_B with a two-sided alternative are given by:

$$H_0: \mu_A - \mu_B = c, H_1: \mu_A - \mu_B \neq c$$

The test statistic is a standard Z -statistic as given by Equation (15.8):

$$Z = \frac{(\bar{Y}_A - \bar{Y}_B) - c}{\sqrt{\sigma_A^2 / n_A + \sigma_B^2 / n_B}} \quad (15.8)$$

If $|Z| > z_{(1-\alpha/2)}$, H_0 is rejected, and we have statistical evidence that $\mu_A - \mu_B \neq c$. For $\alpha = 0.05$, the null hypothesis is rejected if $|Z| > 1.96$. If Z does not exceed 1.96, then H_0 is not rejected.

Example 15.2. Mean percent uranium in UO_2 pellets. A facility producing uranium dioxide (UO_2) pellets wishes to compare the mean percent uranium produced by two different operating processes, A and B . Given population variances and sample sizes $n_A = 8$ and $n_B = 12$, test whether the two process means are the same.

Table 15.3 gives the data and sample statistics.

Table 15.3. A Comparison of two processes for uranium enrichment

Process (treatment)	Process A	Process B
Data	88.056, 88.088	87.939, 87.883, 88.005
	88.044, 88.015	88.064, 88.001, 87.977
	87.897, 88.039	87.881, 87.946, 88.107
	87.950, 88.113	87.970, 87.923, 88.119
Sample size, n_i	8	12
Sample mean \bar{y}_i	88.0252	87.9846
Population variance	0.0071	0.0066

The calculated statistic is:

$$z = \frac{88.0252 - 87.9846}{\sqrt{\frac{0.0071}{8} + \frac{0.0066}{12}}} = 1.073$$

Since $|z|$ is smaller than $z_{0.95} = 1.960$, there is no statistical evidence that the two processes have different means.

- Note that the sample variances are not used in the test because the assumed knowledge of the population variances overrides any estimates. The sample variances, however, may be used for other purposes, such as testing the individual variances or testing whether the population variances are equal. Chapter 14 describes such tests.

For Excel users: Click on Data, Data Analysis (in some versions of Excel, Click on Tools, Data Analysis), and scroll down to “z-test: Two-Sample for Means.” Once a new pane is displayed, select and enter the range of each variable. If the first row contains group labels, check the “Label” box. Otherwise be sure that this box is not checked. Enter the level of significance (default: $\alpha = 0.05$), and the cell where the output begins. You may also enter the constant c (default is $c = 0$) under “Hypothesized Mean Difference.” The output for this procedure gives the calculated t -statistic as well as the critical points for both one- and two-sided tests. Table 15.4 shows Excel’s output, slightly modified.

Table 15.4. Excel output (modified) for Example 15.2

z-Test: Two Samples for Means		
Process	A	B
Mean	88.0230	87.9846
Known variance	0.0071	0.0066
Observations	8	12
Hypothesized mean difference	0	
z	1.073	
Pr{Z<=z}, one-tail	0.142	
z critical one-tail	1.645	
Pr{Z<=z}, two-tail	0.283	
z critical two-tail	1.960	

15.5 Procedure 3: Variances unknown but assumed equal

Procedure 3 applies to perhaps the most common situation that involves a test of a hypothesis about two means. The two populations, designated A and B , are assumed to have equal variances, so that pooling the variances is an important part of the process. The sample sizes are n_A and n_B .

Start by writing the null hypothesis that the group means differ by a constant c , where, typically, $c = 0$:

$$H_0: \mu_A - \mu_B = c$$

Next, write the alternative hypothesis as one of the forms, as appropriate:

$$H_1: \mu_A - \mu_B \neq c, \text{ or } H_1: \mu_A - \mu_B > c, \text{ or } H_1: \mu_A - \mu_B < c$$

Let S^2 be the pooled variance of the two samples, as defined in Section 14.6. The formula for the pooled variance is a special case of Equation (14.10) where $k = 2$. We have:

$$S^2 = \frac{\sum_{i=1}^2 (n_i - 1) S_i^2}{\sum_{i=1}^2 (n_i - 1)} = \frac{(n_A - 1) S_A^2 + (n_B - 1) S_B^2}{n_A + n_B - 2} \quad (15.9)$$

The test statistic is Student's T , given by any one of the three formulas:

$$T = \frac{\bar{Y}_A - \bar{Y}_B - c}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} = \frac{\bar{Y}_A - \bar{Y}_B - c}{\sqrt{S^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} = \frac{\bar{Y}_A - \bar{Y}_B - c}{S \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B} \right)}} \quad (15.10)$$

The calculated t -statistic is then compared to appropriate quantiles of Student's T -distribution with $n_A + n_B - 2$ degrees of freedom from Table T-3 of the appendix.

For a two-sided alternative hypothesis, reject H_0 if $|T| > t_{1-\alpha/2}(n_A + n_B - 2)$.

If the alternative hypothesis is $H_1: \mu_A - \mu_B > 0$, reject H_0 if $T > t_{1-\alpha}(n_A + n_B - 2)$.

If the alternative hypothesis is $H_1: \mu_A - \mu_B < 0$, reject H_0 if $T < -t_{1-\alpha}(n_A + n_B - 2)$.

Example 15.3 illustrates the procedure.

Example 15.3. Measuring radiological contamination. Radiation contamination in a large area was measured, in disintegrations per minute (dpm) per 100 square centimeters (cm^2), in one summer and in the following winter. In each season, a random sample was taken to test whether there is any seasonal variation. The null hypothesis is that the seasonal means are equal (i.e., $c = 0$). We assume that the seasonal variances are equal. Table 15.5 summarizes the collected data.

Table 15.5. Radiological contamination, in dpm/100 cm^2

Sample	Summer	Winter
Mean, \bar{Y}_i	1,988	2,008
Variance, S_i^2	2,051	2,447
Size, n_i	36	39
$\nu_i = n_i - 1$	35	38

Because the population variances are assumed equal, we begin by pooling the sample variances. The pooled variance is calculated as:

$$s^2 = \frac{(35)(2051) + (38)(2447)}{36 + 39 - 2} = 2257.14$$

From Equation (15.10), the denominator of T is:

$$se(\text{mean difference}) = \sqrt{2257.14 \left(\frac{1}{36} + \frac{1}{39} \right)} = 10.98,$$

from which the test statistic is calculated as:

$$t = \frac{(1988 - 2008) - (0)}{10.98} = -1.82$$

The degrees of freedom = $36 + 39 - 2 = 73$. Since $|t|$ is less than $t_{0.975}(73) = 1.99$, there is no statistical evidence that the two seasonal means differ.

For Excel users: Click on Data, Data Analysis (in some versions of Excel, click on Tools, Data Analysis), and scroll down to “*t*-test: Two-Sample Assuming Equal Variances.” Once a new pane is displayed, select and enter the range of each variable. If the first row contains the group labels, check the “Label” box, otherwise be sure that this box is not checked. Then enter the level of significance (default: $\alpha = 0.05$), and the cell where the output begins. Under “Hypothesized Mean Difference,” enter the constant c (default: $c = 0$) against which mean difference is tested. Indicate where the analysis should be printed and Enter “OK.” The output for this procedure gives the calculated *t*-statistic, as well as the critical points for both one- and two-sided tests.

15.6 Procedure 4: Variances unknown and unequal

Procedure 4 is used in the following situation: We have two independent samples from normal distributions with unknown variances σ_A^2 and σ_B^2 that cannot be assumed to be equal as in Procedure 3. We wish to test the null hypothesis $H_0: \mu_A - \mu_B = c$. Instead of using the pooled sample variance S^2 as in Equation (14.10), the test statistic uses the individual sample variances S_A^2 and S_B^2 .

Behrens-Fisher problem

Welch approximation

Satterthwaite approximation

Known as the **Behrens-Fisher problem**, this problem has led to much research since the early decades of the 20th century, and no general exact solution has been found. Brownlee (1965), p. 299, gives a succinct account of the problem—and explains and illustrates a solution in terms of the **Welch approximation** (1937, 1947). The approximate solution we use is the **Satterthwaite approximation** (1946), which extends the Welch approximation from 2 to k groups. The robustness and practicality of this approximation make it an important tool for testing H_0 .

We write the hypotheses and the test statistic as follows:

$$H_0: \mu_A - \mu_B = c, H_1: \mu_A - \mu_B \neq c$$

$$T^* = \frac{(\bar{Y}_A - \bar{Y}_B) - (c)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \quad (15.11)$$

If H_0 is true, T^* has an approximate Student's t -distribution with approximate ν^* degrees of freedom given by:

$$\nu^* = \frac{[S_A^2/n_A + S_B^2/n_B]^2}{\frac{[S_A^2/n_A]^2}{n_A - 1} + \frac{[S_B^2/n_B]^2}{n_B - 1}} \quad (15.12)$$

Because the degrees of freedom is supposed to be an integer, any noninteger value of ν from Equation (15.12) is rounded down. This equation is somewhat unusual; it is one of the few instances in hypothesis testing where the degrees of freedom is determined from the sample data.

Example 15.4. Yield stress of stainless steel pipes. The average yield stress of stainless steel pipes, measured in kilopounds per square inch (ksi), made by two manufacturers is to be compared. We are informed that the variances of the two manufacturers are definitely not equal. Hence, we use Procedure 4. Table 15.6 summarizes data from an experiment conducted at 100 °F.

Table 15.6. Yield stress (in ksi) for two manufacturers of steel pipes

Sample	Manufacturer A	Manufacturer B
Size	5	8
Degrees of freedom	4	7
Mean	82.3	71.4
Variance	108.16	7.84
Standard deviation	10.4	2.8

The hypotheses are:

$$H_0: \mu_A - \mu_B = 0, H_1: \mu_A - \mu_B \neq 0$$

The calculated test statistic is:

$$t^* = \frac{(82.3 - 71.4) - 0}{\sqrt{\frac{108.16}{5} + \frac{7.84}{8}}} = 2.29$$

The calculated degrees of freedom is:

$$v^* = \frac{\left[\frac{108.16}{5} + \frac{7.84}{8} \right]^2}{\frac{\left(\frac{108.16}{5} \right)^2}{4} + \frac{\left(\frac{7.84}{8} \right)^2}{7}} = 4.37$$

This is rounded down to 4 degrees of freedom.

Because this is a double-sided test, the critical value is $t_{0.975}(4) = 2.78$. Since $t^* = 2.29 < 2.79$, we do not reject H_0 . There is insufficient statistical evidence to conclude that the two manufacturers' stainless steel pipes are different with respect to yield stress.

For Excel users. Click on Data, Data Analysis (in some versions of Excel, Click on Tools, Data Analysis), and scroll down to "t-test: Two-Sample Assuming Unequal Variances." Once a new pane is displayed, select and enter the range of each variable. If the first row contains group labels, check the "Label" box; otherwise be sure that this box is not checked. Then enter the level of significance (default: $\alpha = 0.05$), and the hypothesized mean difference (default $c = 0$), and the cell where the output begins. The output for this procedure gives the calculated t -statistic as well as the critical points for α for both one- and two-sided tests. However, the quoted degrees of freedom is rounded and reported to the nearest integer.

15.7 An example to summarize the four procedures

In this section, we have two sets of $n = 10$ observations each. The raw data represent 10 shipments of special material, measured in kilograms, as recorded by both the shipper and the receiver over a period of 1 year. Assuming that all shipments were made by the same shipper and all were received by the same receiver, we know that the proper analysis is that of Procedure 1, paired observations, as given in Section 15.3. In this section, however, we present Excel's printout for each of the four procedures, resulting from different assumptions about the data structure or the assumed variances. We use balanced data (that is, $n_1 = n_2$) because the first procedure requires paired observations, meaning that the number of shipped shipments must equal the number of received shipments. Table 15.7 presents the relevant shipper-receiver data.

Table 15.7. Shipper-receiver differences

Month	Shipper, S_i	Receiver, R_i	Difference, d_i
1	1471.22	1468.12	3.10
2	1470.98	1469.52	1.46
3	1470.82	1469.22	1.60
4	1470.46	1469.26	1.20
5	1469.42	1465.96	3.46
6	1468.98	1470.80	-1.82
7	1469.10	1467.89	1.21
8	1470.22	1472.28	-2.06
9	1470.86	1469.02	1.84
10	1470.38	1470.16	0.22
Mean, \bar{y}			1.021
Std. dev., s			1.818

The setup for the test is:

$$H_0: \mu_d = 0, H_1: \mu_d \neq 0, \alpha = 0.05$$

Excel printouts of four procedures for comparing two means are given in Tables 15.8a through 15.8d.

Table 15.8a. Procedure 1: Paired observations

t-Test: Paired Two-Sample for Means		
	Shipper	Receiver
Mean	1470.244	1469.223
Variance	0.650738	2.942401
Observations	10	10
Pearson Correlation	0.104201	
Hypothesized Mean Difference	0	
Degrees of Freedom	9	
t Stat	1.776051	
P(T<=t) one-tail	0.054728	
t Critical one-tail	1.833113	
P(T<=t) two-tail	0.109456	
t Critical two-tail	2.262157	

Because the t statistic is smaller than $t_{0.95}(9) = 2.26$, we have no evidence of shipper and receiver difference.

Table 15.8b. Procedure 2: Variances known

z-Test: Two-Sample for Means		
	Shipper	Receiver
Mean	1470.244	1469.223
Known Variance	0.75	1.75
Observations	10	10
Hypothesized Mean Difference	0	
Z	2.042	
P(Z<=z) one-tail	0.020576	
z Critical one-tail	1.644854	
P(Z<=z) two-tail	0.041152	
z Critical two-tail	1.959964	

Because the calculated $|z|$ is larger than $z_{(1-\alpha/2)}$, (9) we have statistical evidence of shipper and receiver difference.

Table 15.8c. Procedure 3: Variances unknown but equal

t-Test: Two-Sample Assuming Equal Variances		
	Shipper	Receiver
Mean	1470.244	1469.223
Variance	0.650738	2.942401
Observations	10	10
Pooled Variance	1.796569	
Hypothesized Mean Difference	0	
Degrees of Freedom	18	
t Stat	1.703291	
P(T<=t) one-tail	0.052859	
t Critical one-tail	1.734064	
P(T<=t) two-tail	0.105717	
t Critical two-tail	2.100922	

Because the calculated $|t|$ is smaller than $t_{(1-\alpha/2)}$, (9) we have no statistical evidence of shipper and receiver difference.

Table 15.8d. Procedure 4: Variances unknown and unequal

t-Test: Two-Sample Assuming Unequal Variances		
	Shipper	Receiver
Mean	1470.244	1469.223
Variance	0.650738	2.942401
Observations	10	10
Hypothesized Mean Difference	0	
Degrees of Freedom	13	
t Stat	1.703291	
P(T<=t) one-tail	0.056143	
t Critical one-tail	1.770933	
P(T<=t) two-tail	0.112286	
t Critical two-tail	2.160369	

Because the calculated $|t|$ is smaller than $t_{0.95}(13)$ we have no statistical evidence of shipper and receiver difference.

Note that our calculation of ν^* from equation (15.12) yields $\nu^* = 12.80$. We recommend rounding ν^* down (to 12), Excel rounds ν^* to the nearest integer (to 13).

15.8 Required sample size

As in Sections 13.9 and 13.10, we provide formulas for the required sample size to achieve a specified power at a specified value of the alternative hypothesis for each of the four procedures.

Procedure 1. Paired observations: Because testing paired observations depends only on the differences between the pairs, it is identical to testing a mean with an unknown variance. Accordingly, the procedure in Section 13.10 applies, where n is the number of pairs.

Procedure 2. Variances known: Referring to Section 15.4, we test $H_0: \mu_A - \mu_B = c$ against $H_1: \mu_A - \mu_B \neq c$ at the α level of significance. We assume that independent samples of equal sizes, $n_A = n_B$, are drawn from each population, and we wish to determine the common sample size such that the power is $1 - \beta$ when $\mu_A - \mu_B = c_0$ for $c_0 \neq c$. For testing a one-sided alternative, the required sample sizes are:

$$n_A = n_B = \frac{(z_{1-\alpha} + z_{1-\beta})^2 (\sigma_A^2 + \sigma_B^2)}{(c - c_0)^2} \quad (15.13)$$

Because sample sizes must be integers, round $n_A = n_B$ up to the next largest integer, and use the test statistic from Equation (15.8). Because Equation (15.13) is an approximation to the required sample size, the power requirement is only approximately satisfied.

For testing a two-sided alternative, the required sample sizes are given by Equation (15.13) with $z_{1-\alpha}$ replaced by $z_{1-\alpha/2}$.

Procedure 3. Variances unknown but assumed equal: Referring to Section 15.5, we test $H_0: \mu_A - \mu_B = c$ against $H_1: \mu_A - \mu_B \neq c$ at the α level of significance. As for Procedure 2, we wish to determine a common sample size such that the power is $1 - \beta$ when $\mu_A - \mu_B = c_0$ for $c_0 \neq c$. As in Section 13.10, we use a two-stage test. First, we draw independent random samples of equal sizes, $n_{A1} = n_{B1} = n_1$, from each population and calculate the pooled sample variance S^2 from Equation (15.9). For testing a one-sided alternative, we determine $n_A = n_B = n$ from Equation (15.14) as:

$$n = \frac{2S^2[(t_{1-\alpha}(n_{A1} - 1) + (t_{1-\beta}(n_{B1} - 1))]^2}{(c - c_0)^2} + 0.25[(t_{1-\alpha}(n_{A1} - 1))]^2 \quad (15.14)$$

Because sample sizes must be integers, round n up to the next largest integer. Next, set $n_2 = n - n_1$. If $n_2 < 0$, no additional observations are required. Otherwise, draw n_2 additional observations from each population, calculate the sample means from Equation (13.16), and use the test statistic from Equation (15.10). Because Equation (15.14) is an approximation to the required sample size, the power requirement is only approximately satisfied.

For testing a two-sided alternative, the required sample sizes are given by Equation (15.14) with $t_{1-\alpha}$ replaced by $t_{1-\alpha/2}$.

Procedure 4. Variances unknown and unequal: Referring to Section 15.6, we test $H_0: \mu_A - \mu_B = c$ against $H_1: \mu_A - \mu_B \neq c$ at the $1 - \alpha$ level of significance. As for Procedures 2 and 3, we wish to determine a common sample size such that the power is $1 - \beta$ when $\mu_A - \mu_B = c_0$ for $c_0 \neq c$. This procedure is a modification of Procedure 3 where we use the individual sample variances S_A^2 and S_B^2 instead of the pooled sample variance S^2 . Specifically, we use a two-stage test as in Procedure 3, except that we replace $2S^2$ in Equation (15.14) by $S_A^2 + S_B^2$ to calculate the total sample size n . However, to achieve the power requirement, it may be advisable to increase the sample size to compensate for the inadequacy of the approximation in the replacement for Equation (15.14).

16

One-way ANOVA

16.1 What to look for in Chapter 16

analysis of variance (ANOVA) In Chapter 13, we learned how to test the mean of a normal population. In Chapter 15, we learned how to test the equality of two such means. This chapter provides a methodology, called the *analysis of variance* (often shortened to the acronym *ANOVA*), for testing equality of the means of several normal populations.

ANOVA has many applications. In its simplest application, it allows us to analyze data that are characterized by a single criterion, which is the topic of this chapter. However, even the simplest application of ANOVA involves lengthy calculations. Thus, although we provide the necessary formulas for the construction of the analysis, we recommend that ANOVA calculations be performed and interpreted using appropriate spreadsheet analyses or a full fledged statistical computer software package. Furthermore, we learn to recognize when one-way ANOVA is applicable and when our problem exceeds the scope of this book.

This chapter introduces the following concepts:

- *fixed and random effects, §16.2*
- *balanced design, §16.2*
- *treatments, §16.2*
- *a model for a one-way ANOVA, §16.4*

- total variability and its partition, §16.5
- variability between groups, §16.5
- variability within groups §16.5
- mean square error, §16.6
- *F*-test for one-way ANOVA, §16.6
- interpretation of one-way ANOVA, §16.6
- spreadsheet calculations for one-way ANOVA, §16.7
- Duncan's multiple range test, §16.8

16.2 One-way ANOVA: Data structure

ANOVA analyzes data from several populations or processes and investigates whether the population parameters are different. The methods described here apply to data that are assumed to be normally distributed, although minor departures from normality can generally be tolerated. Data, such as count data, that do not satisfy this assumption may be analyzed using nonparametric methods that do not require normality (see Chapter 25).

Because of its complicated structure, we explain and illustrate one-way ANOVA in the context of an example.

Example 16.1. Enriched uranium from four production lines. We investigate whether four production lines yield the same mean percent of enriched uranium. Table 16.1 shows the data.

Table 16.1. Data layout for percent uranium from four production lines

Group	Production Line 1	Production Line 2	Production Line 3	Production Line 4
$k = 4$ groups	87.2	87.7	88.5	87.5
	87.4	87.7	88.9	87.3
	87.5	88.0	88.5	87.2
$n_1 = 3$		87.8	88.6	87.6
$n_2 = 11$		88.1	88.7	87.4
$n_3 = 15$		87.7	88.7	87.9
$n_4 = 6$		87.6	88.9	
		88.3	88.3	
		88.0	88.8	
		87.8	88.4	
$n_+ = 35$		87.6	88.8	
			88.5	
			88.2	
			88.3	
			88.4	

- The n_+ in the first column in Table 16.1 is $n_1 + n_2 + n_3 + n_4$. This is the plus notation introduced in Chapter 12, where the plus in the subscript indicates the sum over the index of the “plussed” subscripts.

<i>data layout</i>	The data presentation in Table 16.1 is called a <i>data layout</i> because it lays out the data in a fashion that allows us to look at them in a useful way, regardless of their original formatting or the sequence in which they were acquired.
<i>treatments</i>	The four production lines in Table 16.1 correspond to four populations. For ANOVA, these populations are called <i>treatments</i> , and the data from each population is called a <i>treatment group</i> or simply a group. A data layout such as Table 16.1 is called a <i>one-way classification</i> because each observation is classified according to a single classification criterion—its treatment. The data layout shows k treatment groups, where each group is assumed to be a random sample from its respective population. The data analysis for a one-way classification is called a <i>one-way ANOVA</i> .
<i>treatment group</i>	
<i>one-way classification</i>	
<i>one-way ANOVA</i>	
<i>balanced design</i>	If the number of observations in each of the k groups is the same (i.e., $n_1 = \dots = n_k = n$), the experimental design is called a <i>balanced design</i> , and the layout is said to be balanced. Besides being simpler to analyze, balanced designs are desirable because, for a fixed number n_+ of observations, the power of ANOVA is generally higher than with an <i>imbalanced design</i> .
<i>imbalanced design</i>	
<i>fixed effect</i>	When the treatments are completely specified before the data are collected, as is usually the case, the analyses and conclusions apply only to the specified treatments. In this case, the effect of the treatment on the data is called a <i>fixed effect</i> and the associated model for ANOVA is called a <i>fixed effect model</i> .
<i>fixed effect model</i>	
<i>random effect</i>	In contrast, the treatments can be selected at random. For example, suppose there are hundreds of companies competing for a contract to clean up a large contaminated area. Because it is impractical to collect performance data from every company, we may simply investigate whether it makes any difference which company we select. Thus, we may select several companies at random and investigate whether they are significantly different. In this case, the effect of the treatment is called a <i>random effect</i> , and the associated model for ANOVA is called a <i>random effects model</i> .
<i>random effects model</i>	

In either case, the purpose of ANOVA is to investigate whether the treatment effects, whether fixed or random, are statistically different.

The calculations used in the one-way ANOVA of fixed and random effects are the same, but the interpretation and conclusions are different. In more complex ANOVA, the analyses of the two models would be different. In this book, we focus on the fixed effects model, be it one- or multi-way ANOVA.

16.3 Descriptive statistics

The first step is to look at the data. Visualization may tell us whether we have any unexpected problems, such as observations that definitely do not belong with the dataset (i.e., outliers, discussed in Chapter 26). Furthermore, the visual examination may sometimes suggest that normality is unlikely to be a feature of the data or of a specific group within the data. If so, then visualization may suggest a transformation of the data to make it more normal (see Section 16.4). Figure 16.1 presents a graphic display of the data in Example 16.1 and also indicates the group means. Examination of this figure does not reveal any evident problems with the data.

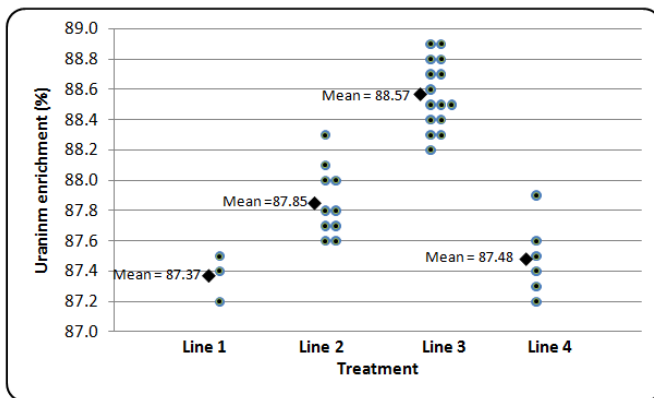


Figure 16.1. Observations plotted by treatment groups

The next step is to calculate the basic descriptive statistics that are used in ANOVA, as shown in Table 16.2.

Table 16.2 Descriptive statistics for Example 16.1

Descriptive statistics	Production Line 1	Production Line 2	Production Line 3	Production Line 4	All Lines
Group subscript: i	1	2	3	4	+
Group size: n_i	3	11	15	6	35
Sum: $y_{i+} = \sum_{j=1}^{n_i} y_{ij}$	262.10	966.30	1328.50	524.90	3081.80
Mean: $\bar{y}_i = \bar{y}_{i+} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$	87.3667	87.8455	88.5667	87.4833	88.0514
Unadjusted sum of squares: $y_{i+}^2 = \sum_{j=1}^{n_i} y_{ij}^2$	22898.85	84885.57	117661.53	46920.31	271366.26
Correction term $\sum_{j=1}^{n_i} y_{ij}^2 / n_i = n_i \bar{y}_i^2$	22898.80	84885.06	117660.82	45920.00	271356.89
(Adjusted) sum of squares: $SS_i = \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{j=1}^{n_i} y_{ij}^2 / n_i$	0.0467	0.5073	0.7133	0.3083	9.3674
Degrees of freedom: $v_i = n_i - 1$	2	10	14	5	34
Variance: $S_i^2 = SS_i / v_i$	0.0233	0.0507	0.0510	0.0617	0.2755

- Note that we obtain the adjusted sum of squares by subtracting the correction term from the unadjusted sum of squares.
- Note that Table 16.2 gives only the descriptive statistics that are needed for the ANOVA. Other descriptive statistics (such as minimum, maximum, range, and standard deviation) are useful for statistical investigations but are not needed for the ANOVA.

16.4 Model and assumptions

Recall that the purpose of the ANOVA is to investigate whether the treatment effects are statistically different. From Figure 16.1, the group means do not all appear to be the same. However, because conclusions from visual inspection are statistically inadmissible, we need a rigorous analysis to determine if the observed differences in the group means are statistically significant. To that end, we begin with a model.

The following model is assumed to describe the data structure for the one-way ANOVA. Let Y_{ij} be the j^{th} observation in the i^{th} treatment group. We have:

$$Y_{ij} = \mu_i + \varepsilon_{ij}; \quad i = 1, \dots, k; \quad j = 1, \dots, n_i \quad (16.1)$$

where:

μ_i = mean of the population associated with the i^{th} group

ε_{ij} = random error associated with the j^{th} observation of the i^{th} group

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

The ANOVA model can also be written in the following equivalent form:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}; \quad i = 1, \dots, k; \quad j = 1, \dots, n_i \quad (16.2)$$

where:

μ = average of the k population means = $(1/k) \sum \mu_i$

$\tau_i = \mu_i - \mu$ = contribution of the i^{th} group = shift of the i^{th} mean away from the average mean

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

The random errors are also assumed to be independent from group to group. Note that the variances of the treatment groups are all assumed equal.

The ANOVA uses a hypothesis test, where the null hypothesis for a fixed effects one-way classification states that the means of all k treatment groups are equal. Symbolically,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (16.3)$$

The alternative hypothesis states that not all the means are equal (i.e., at least two means are different). Symbolically,

$$H_1 : \mu_i \neq \mu_j, \text{ some } i, j \quad (16.4)$$

Note that if all the treatment means are equal, the contribution of every treatment is zero. Therefore, the null hypothesis is equivalent to stating that all τ_i are zero.

As always, the level of significance α must be stated or tacitly understood. By default, $\alpha = 0.05$.

It is essential to clearly understand the assumptions that underlie the ANOVA. If those assumptions are erroneous, so will be our conclusions. If our assumptions are only approximately correct, so will be our conclusions.

We list four assumptions. These assumptions concern the population structure, sample selection, data variability, and distribution of errors.

Assumption 1. The population structure as given by the model is correct. Thus, in Example 16.1, we assume that the four production lines all produce the same uranium quality and that other factors are not involved. For example, we assume that all lines run on the same days of the week, in identical weather conditions, and with supplies coming from the same sources.

Assumption 2. The observations are drawn randomly and independently from one production line to another. Any other procedure runs the risk of introducing a bias into the results.

Assumption 3. The variation about the mean within a group (i.e., the variance) is the same for all groups. As discussed in Chapter 14, this assumption of homoscedasticity can be tested using F_{\max} or Bartlett's test.

Assumption 4. The data are distributed normally. This assumption can be tested using methods discussed in Chapter 11.

ANOVA is a robust procedure, and minor deviations from the last two assumptions are not necessarily serious. Many extensive studies, using real and simulated data, have demonstrated the insensitivity, or robustness, of the analysis to minor departures from these assumptions.

If Assumption 1 is not met, we must reformulate our problem. The following adage may help: It is better to have an approximate solution to an exact problem than an exact solution to an approximate problem.

If Assumption 2 is not met, statistics does not help.

If Assumption 3 is not met, it is sometimes advisable to subdivide the data into clusters of groups that are homogeneous in variance. Often, the lack of homogeneity is the result of extreme values that inflate the variance of one or more groups. In this case, because they are relatively insensitive to extreme observations, nonparametric methods might be useful (see Chapter 25).

If Assumption 4 is not met, two options may be considered. The first is nonparametric analysis, and the second is *data transformation*. The most common transformation is the

logarithmic transformation. We often find that although the observations do not follow the normal distribution, their logarithms (to any base) do. This is often the case with data that have a high degree of skewness. If the data (such as wind velocity) contain zeros, then $\log(Y + c)$ should be used instead of $\log(Y)$. Another transformation that could be tried on the data is the square root transformation. This transformation is often applicable to data with very small values. Further information about transformations may be found in Bartlett (1937).

Discussions of the consequences of incorrect assumptions can be found in papers by Eisenhart et al. (1947) and Cochran (1977).

16.5 Partition of the total sum of squares

Many of the formulas in the next two sections involve lengthy calculations. Fortunately, all these calculations can be made by a spreadsheet routine, as shown in Section 16.7.

As we shall see in Section 16.6, the test statistic for the hypotheses in Equations (16.3) and (16.4) is a function of two estimates of the common variance σ^2 . Each of these estimates is based on a sum of squares, which is the measure of variability used in the ANOVA analysis. Specifically, ANOVA partitions the total data variability into two statistically independent components: variability between the groups and variability within the groups.

Modern communication theory offers a vocabulary that may be *helpful*. Any set of data consists of two components: signal and noise. For ANOVA data, the signal corresponds to the differences in the treatment effects, and the noise corresponds to the statistical variability within the treatment groups. ANOVA attempts to separate the signal from the noise.

The basic relationship for the various sums of squares in ANOVA is:

$$SS_{Total} = SS_{Between} + SS_{Within} \quad (16.5)$$

where:

SS_{Total} = the (adjusted) sum of squares, calculated as if all the data belong to a single group

$SS_{Between}$ = the (adjusted) sum of squares, calculated as a function of the variability of the group means

SS_{Within} = the (adjusted) sum of squares, calculated as a function of the population variability pooled across all groups

The formula for calculating SS_{Total} is given by:

$$SS_{Total} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 \quad (16.6)$$

or by the equivalent working formula:

$$SS_{Total} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - CT \quad (16.7)$$

where CT is the correction term, calculated as:

$$CT = \frac{1}{n_+} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} \right)^2 = n_+ \bar{Y}^2 \quad (16.8)$$

and where:

$$\bar{Y} = \frac{1}{n_+} \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} \quad (16.9)$$

is the average of all n_+ observations.

The degrees of freedom associated with SS_{Total} is:

$$v_{Total} = n_+ - 1 \quad (16.10)$$

In Example 16.1, SS_{Total} is calculated from Table 16.2 and Equation (16.7) as:

$$SS_{Total} = 271366.26 - 35(88.0514)^2 = 9.3674, \text{ with } v_{Total} = 35 - 1 = 34$$

- Note that hand-held calculators may yield different results here and in other calculations. These discrepancies depend on the number of significant figures that are used throughout the calculations, including intermediate results. The calculations reported here agree with the Excel calculations, shown in Table 16.6. $SS_{Between}$ is calculated either by the definition formula (Equation 16.11) or by the equivalent working formula (Equation 16.12).

$$SS_{Between} = \sum_{i=1}^k \sum_{j=1}^{n_i} \bar{Y}_i - \bar{Y}^2 = \sum_{i=1}^k n_i \bar{Y}_i - \bar{Y}^2 \quad (16.11)$$

$$SS_{Between} = \sum_{i=1}^k CT_i - CT = \sum_{i=1}^k n_i \bar{Y}_i^2 - n_+ \bar{Y}^2 \quad (16.12)$$

where CT is the correction term from Equation (16.8) and CT_i is the correction term for the i^{th} group, calculated as:

$$CT_i = \frac{1}{n_i} \left(\sum_{j=1}^{n_i} Y_{ij} \right)^2 = n_i \bar{Y}_i^2 \quad (16.13)$$

The degrees of freedom associated with $SS_{Between}$ is:

$$v_{Between} = k - 1 \quad (16.14)$$

In Example 16.1, $SS_{Between}$ is calculated from Table 16.2 and Equation (16.12) as:

$$SS_{Between} = 22898.80 + \dots + 45920.00 - 35(88.0514)^2 = 7.7918, \text{ with}$$

$$v_{Between} = 4 - 1 = 3.$$

➤ $SS_{Between}$ is sometimes called *among groups sum of squares* and is written SS_{Among} . Some texts prefer the notation SS_{Group} or $SS_{Treatment}$.

SS_{Within} is calculated by either of two methods. The first method is to add the sums of squares of the individual groups. SS_{Within} is calculated by the definition formula:

$$SS_{Within} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \quad (16.15)$$

or by the equivalent working formula:

$$SS_{Within} = \sum_{i=1}^k \left[\sum_{j=1}^{n_i} Y_{ij}^2 - n_i \bar{Y}_i^2 \right] = \sum_{i=1}^k \text{Adjusted sum of squares of group } i \quad (16.16)$$

In Example 16.1, SS_{Within} is calculated from Table 16.2 and Equation (16.16) as:

$$SS_{Within} = 0.0467 + 0.5073 + 0.7133 + 0.3083 = 1.5756$$

The second, and shorter, method to calculate SS_{Within} is to use subtraction. From Equation (16.5):

$$SS_{Within} = SS_{Total} - SS_{Between} \tag{16.17}$$

In Example 16.1, using this formula and the calculated values of SS_{Total} and $SS_{Between}$, we have $SS_{Within} = 9.37 - 7.79 = 1.58$, which agrees with the value calculated using the first method.

The degrees of freedom associated with SS_{Within} is:

$$v_{Within} = n_+ - k \tag{16.18}$$

The latter may also be obtained by subtraction as:

$$v_{Within} = v_{Total} - v_{Between} = (n_+ - 1) - (k - 1) = n_+ - k \tag{16.19}$$

In Example 16.1, the associated degrees of freedom for SS_{Within} is:

$$v_{Within} = 35 - 4 = 31$$

16.6 The one-way ANOVA table

Table 16.3 presents the general one-way ANOVA table showing all its elements. The new elements are the last two columns: mean square and the F ratio. The mean squares are estimators of the common variance σ^2 , and F is the test statistic.

Table 16.3. General one-way ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Between groups	$SS_{Between}$	$v_{Between}$	$MS_{Between} = \frac{SS_{Between}}{v_{Between}}$	$F = \frac{MS_{Between}}{MS_{Within}}$
Within groups	SS_{Within}	v_{Within}	$MS_{Within} = \frac{SS_{Within}}{v_{Within}}$	
Total	SS_{Total}	v_{Total}		

between groups mean square The first mean square, called the **between groups mean square**, is denoted by $MS_{Between}$ and is defined as:

$$MS_{Between} = SS_{Between} / v_{Between} \tag{16.20}$$

If the null hypothesis of equal means is correct, $MS_{Between}$ is an unbiased estimator of σ^2 .

In Example 16.1, $MS_{Between}$ is calculated as $7.79/3 = 2.5973$.

within groups mean square The second mean square, called the **within groups mean square**, is denoted by MS_{Within} and is defined as:

$$MS_{Within} = SS_{Within}/v_{Within} \quad (16.21)$$

Because SS_{Within} is based on the means of the various groups (see Equation (16.15)), MS_{Within} is an unbiased estimator of σ^2 regardless of the correctness of the null hypothesis.

In Example 16.1, MS_{Within} is calculated as $1.5756 / 31 = 0.0508$.
The ratio

$$F = MS_{Between} / MS_{Within} \quad (16.22)$$

F ratio variance ratio is the last entry in the ANOVA table. The **F ratio**, sometimes called the **variance ratio**, is the test statistic. If the null hypothesis is correct, this statistic is distributed as an F variable with $v_{Between}$ degrees of freedom in the numerator and v_{Within} degrees of freedom in the denominator.

In Example 16.1, the F statistic is calculated as $2.5973/0.0508 = 51.10$.

If H_0 is not correct and not all the means are equal, then the sums of squares in the numerator $MS_{Between}$ of F will have a component depending on the differences in the means and will tend to be larger than in a case where H_0 is true. Hence, the test is one-sided, and H_0 is rejected if F exceeds the critical value. The critical value is $F_{1-\alpha}(v_{Between}, v_{Within})$ from Table T-4 of the appendix. If the null hypothesis of equality of means is rejected, we conclude that there is statistical evidence that the group means are not all equal. If H_0 is not rejected, there is no statistical evidence that the groups are different.

- If H_0 is not true, the distribution of F will tend to be larger than if H_0 is true. Hence, if the calculated value of F corresponds to a very low quantile of the F -distribution, we may suspect data tampering.

For Example 16.1, Table 16.4 shows the calculated ANOVA table.

Table 16.4. ANOVA table for Example 16.1

Source of variation	Sum of squares	Degrees of freedom	Mean squares	<i>F</i>
Between groups	7.79	3	2.60	51.1
Within groups	1.58	31	0.05	
Total	9.37	34		

In Example 16.1, using $\alpha = 0.05$, we have $F_{0.95}(3, 31) = 2.91$. Since $51.1 > 2.91$, we reject the hypothesis of equal means, citing statistical evidence that not all production lines have the same average.

16.7 Constructing an ANOVA Table with Excel

Excel is a convenient tool for the construction of an ANOVA table. Under “Data” (in some versions of Excel under “Tools”), select “Data Analysis,” and then select “ANOVA: Single Factor.” In the ANOVA panel, select the data array. If the data are selected with labels as column headers, check the box “Labels in First Row.” Next, enter the level of significance (defaults to 0.05), and indicate the cell where the printout should start.

Two tables show Excel’s output for the data in Example 16.1. Table 16.5 displays some summary (descriptive) statistics, while Table 16.6 displays the ANOVA table. Note that Excel’s ANOVA also provides the critical value for the analysis, obviating the need for table lookup and interpolation.

Table 16.5. Excel’s summary statistics for Example 16.1

Group	Count	Sum	Average	Variance
Line 1	3	262.1	87.3667	0.0233
Line 2	11	966.3	87.8455	0.0507
Line 3	15	1328.5	88.5667	0.0510
Line 4	6	524.9	87.4833	0.0617

The Excel printout continues with the one-way ANOVA (called “Single Factor” by Excel).

Table 16.6. Excel's ANOVA table for Example 16.1

Source of Variation	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	7.7918	3	2.5973	51.1013	4.18E-12	2.911
Within Groups	1.5756	31	0.0508			
Total	9.3674	34				

Because Excel calculates the ANOVA table from the raw data, it cannot construct the table from the intermediate descriptive statistics. There is no Excel function or routine that constructs an ANOVA table from descriptive statistics such as those given in Table 16.2.

- Excel gives the critical point for the selected α (here, $\alpha = 0.05$) in the column labeled “F crit.” In addition, Excel also gives the *P*-value, which is the probability of obtaining an *F*-value at least as large as the one that was actually observed (in our example, 51.1013), assuming that the null hypothesis is true. That probability is reported as 4.18×10^{-12} , which is certainly smaller than $\alpha = 0.05$.

16.8 Duncan's multiple range test

Suppose we perform an ANOVA and reject the hypothesis of equal means. If we have more than two means, a natural question is how the means are different. Is just one of the means very much different from the others? Are there several clusters of means? Or, is each mean different from all of the other means?

We may be tempted to separately test all possible pairs of means. For three groups, this would involve three tests: Group *A* versus Group *B*, Group *A* versus Group *C*, and Group *B* versus Group *C*. However, with more than two groups, the probability of a Type I error will be much larger than the nominal value of α used for each of the tests. This happens because the probability of having at least one statistically significant difference increases with the number of comparisons. For example, if four groups are tested in a pair-wise fashion, each with $\alpha = 0.05$, there is about a 25% chance that at least one of the six tests will incorrectly reject H_0 if all the means are equal. For a discussion of this problem, see Dixon and Massey (1983), p. 131.

multiple range tests A widely used solution to this problem employs one of the many techniques that are called **multiple range tests** or **multiple comparisons**. These techniques are designed to identify which means are different at a specified probability of Type I error.

multiple comparisons Caution is needed in choosing one of the many multiple range tests available, because applying different multiple range tests to the same set of multigroup data does not necessarily yield identical results. The various multiple range tests are constructed under somewhat different assumptions and provide different degrees of conservatism. A presentation showing the full variety of multiple range testing is beyond the scope of this text. See Toothaker (1991) for an extended discussion of this knotty, often perplexing, and sometimes controversial problem.

Duncan's multiple range test However, despite these drawbacks, multiple range tests provide an important decisionmaking tool. One of the most popular—and, at the same time, most useful and intuitive—is **Duncan's multiple range test** (Duncan, 1986) See Steel and Torrie (1980), Chapter 8, for a thorough discussion. Bowen and Bennett (1988), p. 256, offer an exposition of Duncan's multiple range test; Example 16.2 lays out their approach in the following continuation of Example 16.1.

Example 16.2. Duncan's multiple range test. For convenience, we assemble the relevant values and results from the analysis of Example 16.1.

Experiment size: $n_+ = 35$

Number of groups: $k = 4$

Group sizes: $n_1 = 3, n_2 = 11, n_3 = 15, n_4 = 6$

Group means: $\bar{y}_1 = 87.37, \bar{y}_2 = 87.85, \bar{y}_3 = 88.57, \bar{y}_4 = 87.48$

Within groups mean square (from Table 16.6): $MS_{Within} = 0.0508$

Now proceed with the following steps:

harmonic mean **Step 1.** Calculate \bar{n}_h , the **harmonic mean** of the k sample sizes:

$$\bar{n}_h = k / (1/n_1 + 1/n_2 + \dots + 1/n_k) \quad (16.23)$$

➤ If all groups are of the same size (i.e., $n_1 = n_2 = \dots = n_k = n$), then $\bar{n}_h = n$.

In Example 16.1 we calculate:

$$\bar{n}_h = 4 / (1/3 + 1/11 + 1/15 + 1/6) = 6.0829$$

Step 2. Calculate a pseudo standard deviation for the sample means:

$$S_{\bar{Y}}^* = \sqrt{\frac{MS_{Within}}{n_h}} \quad (16.24)$$

For Example 16.1, we have:

$$S_{\bar{y}}^* = \sqrt{\frac{0.0508}{6.0829}} = 0.0914$$

Step 3. From Table T-7 of the appendix, find the values of $q_{0.95}(p, n_+ - k)$ for $p = 2, 3, \dots, k$. Table interpolation may be needed.

➤ Table T-7 provides these values only for $\alpha = 0.05$. Table values for other levels of significance may be found in Duncan (1986).

In Example 16.1, we have $n_+ - k = 35 - 1 = 31$. For $\alpha = 0.05$, the following table values are found using interpolation:

$$q_{0.95}(2, 31) = 2.89$$

$$q_{0.95}(3, 31) = 3.04$$

$$q_{0.95}(4, 31) = 3.12$$

Step 4. Calculate R_p^* for $p = 2, 3, \dots, k$. These are the least significant ranges and will be the critical values for Duncan's multiple range test.

$$R_p^* = [q_{(1-\alpha)}(p, N - k)] S_{\bar{Y}_i}^* \quad (16.25)$$

For Example 16.1, we have:

$$R_2^* = 2.89(0.0932) = 0.27$$

$$R_3^* = 3.04(0.0932) = 0.28$$

$$R_4^* = 3.12(0.0932) = 0.29$$

Step 5. Arrange the sample means in ascending order. For Example 16.1, the ordered means are displayed below. For clarity, the spacing between the means is schematically proportioned to the numerical differences between successive means.

$$\begin{array}{ccccccc} \bar{y}_1 & & \bar{y}_4 & & \bar{y}_2 & & \bar{y}_3 \\ 87.37 & & 87.48 & & 87.85 & & 88.57 \end{array}$$

Step 6. Denote the rank of the smallest sample mean by 1, the next smallest by 2, ..., and the largest by k . For any two means, let p denote the difference between the corresponding ranks plus 1. Thus, for the largest and the smallest of k means, $p = (k - 1) + 1 = k$.

Step 7. The ranges (or differences) of all pairs of sample means are tested for equality in $k - 1$ stages. The procedure for doing this is illustrated for Example 16.1, where $k = 4$

The first stage tests the largest sample mean against each of the $k - 1 = 3$ smaller sample means by comparing the ranges of the $(k - 1)$ pairs to their corresponding critical values from Step 4. The first stage comparisons are as follows:

$$\begin{aligned}\bar{y}_3 - \bar{y}_1 &= 88.57 - 87.37 = 1.20 > R_4^* = 0.29 \\ \bar{y}_3 - \bar{y}_4 &= 88.57 - 87.48 = 1.09 > R_3^* = 0.28 \\ \bar{y}_3 - \bar{y}_2 &= 88.57 - 87.85 = 0.72 > R_2^* = 0.27\end{aligned}$$

Because all three ranges are larger than their critical values, we conclude that \bar{y}_3 is significantly larger than any of the other smaller sample means.

The second stage continues by testing the second largest sample mean against each of the $k - 2 = 2$ smaller sample means as in the first stage. The second stage comparisons are as follows:

$$\begin{aligned}\bar{y}_2 - \bar{y}_1 &= 87.85 - 87.37 = 0.48 > R_3^* = 0.28 \\ \bar{y}_2 - \bar{y}_4 &= 87.85 - 87.48 = 0.37 > R_2^* = 0.27\end{aligned}$$

Because both ranges are larger than their critical values, we conclude that \bar{y}_2 is significantly larger than either \bar{y}_1 or \bar{y}_4 .

The third and, for this example, the last stage tests the second smallest against the smallest sample mean. The last stage comparison is:

$$\bar{y}_4 - \bar{y}_1 = 87.48 - 87.37 = 0.11 < R_2^* = 0.27$$

We conclude that \bar{y}_4 and \bar{y}_1 are not significantly different.

The result of each stage in Step 7 is a set of conclusions that the differences between selected pairs of sample means are or are not statistically significant. However, it sometimes happens that two sample means that are significantly different fall between two means that are not significantly different. To avoid contradictions, a difference between two sample means is considered non-significant if both means fall between two other means that do not differ

significantly. (The same conclusion holds if one of the first two means coincides with one of the non-significant means.) With this convention, the test results are a list of the ordered sample means grouped into non-overlapping clusters, with all means in a cluster not significantly different and means in different clusters significantly different.

A graphical display, described by Duncan (1986), p. 757, is useful for describing multiple range test results. Write the means in an increasing sequence, left-to-right, and underline those collections of means that are *not* significant by the multiple range test. For Example 16.1, we have this display:

\bar{y}_1	\bar{y}_4	\bar{y}_2	\bar{y}_3
87.37	87.48	87.85	88.57

This display indicates that only the means of Line 4 and Line 1 are not statistically significantly different (i.e., they form a cluster), as do each of the means of Line 2 and Line 3.

16.9 T-test and ANOVA equivalence

The one-way ANOVA tests the equality of two or more means. However, in Chapter 15, we developed a *t*-test for comparing two means. The natural question is: Are the *t*-test and the ANOVA for two groups equivalent? The answer is a definite yes, provided, that the tests are based on the same assumptions.

We are given random samples from two normal populations with the same unknown variance and wish to test equality of the means, where the null and the alternative hypotheses are:

$$H_0: \mu_1 = \mu_2, \text{ and } H_1: \mu_1 \neq \mu_2 \quad (16.26)$$

The *t*-test is described in Procedure 3 of Chapter 15 with test statistic $T = T(\nu)$ from Equation (15.10), where $\nu = n_A + n_B - 2$ is the degrees of freedom for the pooled variances and the test is two-sided. In the corresponding ANOVA, we have $k = 2$ groups and use the test statistic $F = F(\nu_{\text{Between}}, \nu_{\text{Within}})$ from Equation (16.22), where the degrees of freedom in the numerator and denominator of the *F* statistic are $k - 1 = 1$ and $n_+ - 2$, respectively. To begin, the total number of observations is $n_A + n_B = n_+$, so that the degrees of freedom ν and ν_{Within} are equal. Continuing, it can be shown that:

$$(T(\nu))^2 = F(1, \nu) \quad (16.27)$$

In other words, the square of a *t*-statistic with ν degrees of freedom is equal to an *F*-statistic with one degree of freedom in the numerator and ν degrees of

freedom in the denominator. Because the t -distribution is symmetric about zero and the F -distribution is positive, a tail probability for F is twice the corresponding right-tail probability for t . It follows from Equation (16.27) that, for any α ,

$$\left(t_{1-\alpha/2}(\nu)\right)^2 = F_{1-\alpha}(1, \nu) \quad (16.28)$$

Because the quantiles in Equation (16.28) are the critical values for the two-sided t - and F -tests, these two tests will always give the same results when applied to the same data. Therefore, the ANOVA for two groups is equivalent to the t -test. As an example, for $\alpha = 0.10$ and $\nu = 9$, from Tables T-3 and T-4 of the appendix, we have:

$$\left(t_{0.95}(9)\right)^2 = 1.833113^2 = 3.360303 \text{ and } F_{0.90}(9) = 3.360303.$$

- Note that the tables in the appendix do not have enough significant figures for an exact comparison. We used Excel for this verification.

17

Two-way ANOVA

17.1 What to look for in Chapter 17

A natural extension of one-way analysis of variance (ANOVA) is two-way ANOVA, the subject of this chapter. In a two-way ANOVA, we test the effects of each of two factors and their combination on the response variables (i.e., observations). As in Chapter 16, the observations are assumed to be distributed normally. When normality is not applicable, we may use a nonparametric analysis, as described in Chapter 25.

This chapter discusses the following concepts and terms, some of which we have encountered previously:

- *factorial design*, §17.2
- *factor*, §17.2
- *level*, §17.2
- *cross-classification*, §17.2
- *main effect*, §17.2
- *interaction*, §17.2
- *experimental unit*, §17.2
- *replication*, §17.2
- *block*, §17.3

- *randomized complete block design*, §17.3
- *fixed, random, and mixed effects*, §17.4
- *two-way ANOVA without replication*, §17.5, §17.6
- *two-way ANOVA with replication* §17.7

17.2 Two-way factorial designs

<i>factorial design</i>	The one-way ANOVA introduced in Chapter 16 investigates the effect of a single factor (i.e., the treatment) on the observations. A two-way factorial design is characterized by two classification criteria called factors , denoted here by Factor <i>A</i> and Factor <i>B</i> .
<i>factor</i>	(Although there are <i>N</i> -way factorial designs, this book discusses only two-way factorial designs, which will be referred to simply as factorial designs.) Each factor has two or more levels , with Factor <i>A</i> having <i>a</i> levels and Factor <i>B</i> having <i>b</i> levels. This data classification is analogous to the structure of a contingency table discussed in Chapter 12. A factorial design is represented by an $a \times b$ data array, where rows refer to the levels of Factor <i>A</i> , columns refer to the levels of Factor <i>B</i> , and cell (i, j) refers to the intersection of the i^{th} row and the j^{th} column. The entry in cell (i, j) includes all of the observations on the response variable at the corresponding factor levels. A complete factorial design includes at least one observation at each of the $a \times b$ level combinations.
<i>level</i>	
<i>complete factorial design</i>	
<i>cross-classification</i>	A more descriptive name for factorial design is cross-classification design.
<i>response variable</i>	As for the one-way ANOVA, the observations are denoted by <i>Y</i> , which is the dependent variable or response variable . The effect of a factor is defined as the change in <i>Y</i> associated with a change in the level of the factor. The effect of any one factor on <i>Y</i> is called a main effect . The combined (or joint) effect of both factors separate from the main effects is called an interaction effect or simply an interaction . If the main effects completely explain the effects of the factors on <i>Y</i> , we say that there is no interaction.
<i>effect</i>	
<i>main effect</i>	
<i>interaction</i>	

For example, the effect of temperature (Factor *A*) without regard to pressure on the yield of a chemical process is a main effect, and so is the effect of pressure (Factor *B*) without regard to temperature. The interaction is the effect on the process yield that is not explained by the separate main effects (i.e., temperature and pressure).

<i>experimental unit</i>	An item or individual portion of material measured or tested during an experiment is called an <i>experimental unit</i> . For example, if we measure the permeability of concrete by applying a “scratch test” to 1-square-meter areas, each such area is an experimental unit. The investigation or measurement of a single experimental unit is called a <i>run</i> or a <i>trial</i> .
<i>run</i>	
<i>trial</i>	
<i>replication</i>	A repetition of all or part of an experiment is called a <i>replication</i> . For example, if we weigh each shipment twice before it leaves the facility, we have two replications. Replications allow us to estimate the experimental error variance, which is the variance of experimental units that are treated alike. If the number of replications is the same for every experimental unit, we have a <i>balanced design</i> . The calculations involved in the analysis of a balanced design are simpler than those for an unbalanced design. Also, spreadsheets such as Excel do not include ANOVA routines for unbalanced designs, other than for a one-way ANOVA.
<i>balanced design</i>	

This chapter is restricted to two-way factorial designs. The analysis of three-way or higher factorial designs is complicated and requires specialized statistical software.

17.3 Randomized complete block design

<i>randomized complete block design</i>	A <i>randomized complete block design</i> is a special case of a complete factorial design. The essence of a randomized complete block design is that the experimental units are divided into homogeneous groups called <i>blocks</i> that correspond to different levels of one factor, say, Factor <i>A</i> . The experimental units within a block are randomly assigned to the levels of the second factor, say, Factor <i>B</i> , the factor being investigated. In the absence of Factor <i>B</i> , observations within a block are expected to be alike, but the observations might differ from block to block. Blocks play the same role in a randomized complete block design as do pairs in a paired observation design (Section 15.3). The random assignment of treatments to blocks tends to reduce the experimental error and hence increase the sensitivity of the analysis to real differences among the various levels of Factor <i>B</i> .
<i>block</i>	

To illustrate this design, consider a farming experiment. Suppose a farmer wishes to compare four varieties of apples for yield (total fruit weight per tree). The farmer randomly selects three sections of land throughout his farm, divides each section into four plots of equal size, randomly assigns the four varieties of apples to the four plots, and plants the same number of trees in each plot. The three blocks are the sections of land, and Factor *A* has three levels. The experimental units are the plots, and Factor *B* has four levels, one for each apple

variety. In addition to being the same size and having the same number of trees, all plots receive similar amounts of water and fertilizer. At harvest time, the farmer weighs the apple yield of each tree in each plot. Note that, if the four varieties are supposed to be alike, we expect all trees within the same plot to have similar yields.

In statistical terms, the yield is the dependent or response variable, denoted by Y . The three sections of land are the three blocks or levels of Factor A . The four apple varieties are the four levels of Factor B . The plots are the experimental units. Multiple trees within the same plot are replications.

Example 17.1. Ice sublimation. In an ice-condenser reactor, ice baskets are in arrays called sectors. Each sector contains the same number of columns, and each column contains the same number of rows, where the first row is the closest to and the last row is farthest away from the reactor. Because the position of the column is considered to have no effect, each column is a block.

At the beginning of a year all ice baskets are assumed to have the same known weight. Over the course of the year, the ice loses weight because of sublimation (i.e., evaporation). At the end of the year, some baskets are weighed in order to estimate the average weight loss and to assess whether there is a row effect (i.e., whether proximity to the reactor affects weight loss).

<i>fixed effects</i>	If a study is designed so that only selected rows and columns are involved, the design is a <i>fixed effects</i> design, and the conclusions from the data analysis apply only to the selected rows and columns.
<i>random effects</i>	If both rows and columns are selected at random, the design is a <i>random effects</i> design, and the conclusions apply to the entire array.
<i>mixed effects</i>	If the columns are selected at random but the rows are specifically selected (or the other way around), the design is a <i>mixed effects</i> design and the conclusions apply only to the selected rows.

17.4 Data structure and model: No replication

Table 17.1 shows the data structure for a complete two-way factorial design with no replication. Factor B might be a block factor or a treatment. Section 17.7 considers factorial designs with replications

Table 17.1. Data structure for a complete two-way factorial design

Factor A levels	Factor B levels			
	1	2	...	<i>b</i>
1	y_{11}	y_{12}	...	y_{1b}
2	y_{21}	y_{22}	...	y_{2b}
⋮	⋮	⋮	⋮	⋮
<i>a</i>	y_{a1}	y_{a2}	...	y_{ab}

Example 17.1 (continued). Ice sublimation. The plant management decides to conduct a pilot study to investigate whether proximity to the reactor affects sublimation. Six ice baskets are selected from two columns and three rows, and their weight losses (in kilograms (kg)) are measured. The columns selected are the two edge columns of the horseshoe, and from each column, an ice basket is selected from each of three rows: first, middle, and last. Table 17.2 shows the data. The rows are the Factor A levels, and the columns are the Factor B levels, which are treated as blocks. Table 17.2 also shows the intermediate calculations that are needed to construct the various sums of squares in the ANOVA.

Table 17.2. Ice weight loss (kg) in 1 year

Factor A	---- Factor B ----		Sum	Mean	SS, unadjusted	Correction Term	SS, adjusted
	Column 1	Column 2					
First row	50	54	104.0	52.0	5416.0	5408.0	8.0
Middle row	22	26	48.0	24.0	1160.0	1152.0	8.0
Last row	24	25	49.0	24.5	1201.0	1200.5	0.5
Sum	96.0	105.0	201.0				
Mean	32.0	35.0		33.5			
SS, unadjusted	3560.0	4217.0			7777.0		
Correction Term	3072.0	3675.0				6733.5	
SS, adjusted	488.0	542.0					1043.5

Corresponding to Table 17.1, let Y_{ij} be the observation at the i^{th} level of Factor A and the j^{th} level of Factor B in cell (i, j) . As for a one-way ANOVA, the model

for a complete two-way factorial design with no replication is an additive model. We write:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, b \quad (17.1)$$

where:

μ = overall mean response

α_i = main effect of the i^{th} level of Factor A averaged over the levels of Factor B

β_j = main effect of the j^{th} level of Factor B averaged over the levels of Factor A

ε_{ij} = random error associated with Y_{ij}

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

For a fixed effects model, the following holds:

$$\sum \alpha_i = 0 \text{ and } \sum \beta_j = 0 \quad (17.2)$$

For a random effects model, the factor levels are assumed to be random variables such that:

$$\alpha_i \sim N(0, \sigma_A^2) \text{ and } \beta_j \sim N(0, \sigma_B^2) \quad (17.3)$$

where σ_A^2 and σ_B^2 are the variances of Factor A and Factor B effects, respectively.

For a mixed effects model, only one of the following holds:

$$\sum \alpha_i = 0 \text{ and } \beta_j \sim N(0, \sigma_B^2) \quad (17.4)$$

$$\sum \beta_j = 0 \text{ and } \alpha_i \sim N(0, \sigma_A^2)$$

Two null hypotheses are tested simultaneously. For a fixed effects model, we test whether the factor levels are the same. The hypotheses for Factor A are:

$$H_0: \alpha_i = 0, \text{ all } i \text{ vs. } H_1: \alpha_i \neq 0, \text{ some } i \quad (17.5)$$

Since $\sum \alpha_i = 0$, this test is equivalent to:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a \text{ vs. } H_1: \alpha_i \neq \alpha_j, \text{ some } i, j \quad (17.6)$$

The hypotheses for Factor B are:

$$H_0: \beta_j = 0, \text{ all } j \text{ vs. } H_1: \beta_j \neq 0, \text{ some } j \quad (17.7)$$

Since $\sum \beta_j = 0$, this test is equivalent to testing:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_b \text{ vs. } H_1: \beta_i \neq \beta_j, \text{ some } i, j \quad (17.8)$$

For a random effects model, we test whether the factor variances are zero. The hypotheses tests are:

$$H_0: \sigma_A^2 = 0 \text{ vs. } H_1: \sigma_A^2 \neq 0 \text{ and} \quad (17.9)$$

$$H_0: \sigma_B^2 = 0 \text{ vs. } H_1: \sigma_B^2 \neq 0 \quad (17.10)$$

For a mixed effects model, where Factor A is a fixed effect and Factor B is a random effect, the hypotheses are:

$$H_0: \alpha_i = 0, \text{ all } i \text{ vs. } H_1: \alpha_i \neq 0, \text{ some } i \quad (17.11)$$

and

$$H_0: \sigma_B^2 = 0 \text{ vs. } H_1: \sigma_B^2 \neq 0 \quad (17.12)$$

The assumptions made in the analysis of two-way factorial designs are similar to those made for a one-way design in Section 16.4. They concern the population structure, sample selection, data variability, and distribution of errors:

Assumption 1. The population structure as given by the model is correct.

Assumption 2. For random effects models, the experimental units are drawn randomly and independently.

Assumption 3. The variance of the responses is the same for all cells.

Assumption 4. The data are distributed normally.

17.5 ANOVA for a two-way factorial design without replication

The ANOVA for a two-way factorial design is an extension of the one-way ANOVA in Chapter 16. The ANOVA table has four lines, representing four sources of variation. Each source of variation has an associated sum of squares (SS) and degrees of freedom.

The sum of squares for total variability is partitioned as follows:

$$SS_{Total} = SS_A + SS_B + SS_{Error} \quad (17.13)$$

where:

SS_{Total} = the (adjusted) sum of squares for the variability of the data about the overall mean

SS_A = the (adjusted) sum of squares for the variability of Factor A 's group means

SS_B = the (adjusted) sum of squares for the variability of Factor B 's group means

SS_{Error} = the (adjusted) sum of squares for error

➤ SS_{Error} is sometimes called $SS_{Residual}$ because it is the part of SS_{Total} that is left over after accounting for SS_A and SS_B .

Equation (17.14) gives the degrees of freedom associated with the sums of squares.

$$\begin{aligned} v_{Total} &= n_+ - 1 \\ v_A &= a - 1 \\ v_B &= b - 1 \\ v_{Error} &= (a - 1)(b - 1) \end{aligned} \quad (17.14)$$

where $n_+ = ab =$ total number of observations

Table 17.3 displays the general ANOVA table with all its elements. The mean squares are the sums of squares divided by their corresponding degrees of freedom. The two F ratios are the mean squares divided by the mean square for error.

Note the similarity between Table 17.3 and Table 16.3. Both tables show the total variability partitioned into independent components. Both tables show an

estimate of σ^2 . This estimate is the mean square of the “within group” in Table 16.3 and “Error” in Table 17.3. The difference between the two tables is that the “between groups” sources of variability of Table 16.3 is split in Table 17.3 between two sources of variability that (in the context of example 17.1) are called “Rows” and “Columns.”

Table 17.3. ANOVA table for a two-way factorial design

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Factor A	SS_A	ν_A	$MS_A = \frac{SS_A}{\nu_A}$	$F_A = \frac{MS_A}{MS_{Error}}$
Factor B	SS_B	ν_B	$MS_B = \frac{SS_B}{\nu_B}$	$F_B = \frac{MS_B}{MS_{Error}}$
Error	SS_{Error}	ν_{Error}	$MS_{Error} = \frac{SS_{Error}}{\nu_{Error}}$	
Total	SS_{Total}	ν_{Total}		

- The mean square for error is the estimator of the variance σ^2 of ε_{ij} in Equation (17.1).

To test the hypotheses stated in Section 17.4, compare each of the two F -statistics to an appropriate critical value. For Factor A, the critical value is $F_{1-\alpha}(\nu_A, \nu_{Error})$, and for Factor B, the critical value is $F_{1-\alpha}(\nu_B, \nu_{Error})$. If an F -statistic exceeds its critical value, the null hypothesis for the associated factor is rejected.

The same ANOVA table and critical values are used to test hypotheses for all two-way factorial experiments with a single observation (no replication) per experimental unit. Regardless of the design (randomized complete block or two-way factorial), or the model type (fixed effects, random effects, or mixed effects), all ANOVA calculations are the same. The only difference in the analysis is in the interpretation of the results, as outlined at the end of Section 17.3.

We return to Example 17.1. Because this is a fixed effects design, the hypotheses are given by Equations (17.6) and (17.8):

$$H_0: \alpha_{First} = \alpha_{Middle} = \alpha_{Last} \text{ vs. } H_1: \alpha_i \neq \alpha_j, \text{ some } i, j$$

$$H_0: \beta_{Column 1} = \beta_{Column 2} \text{ vs. } H_1: \beta_{Column 1} \neq \beta_{Column 2}$$

The calculations of the various sums of squares and mean squares for example 17.1 are not shown here. These calculations, required for the ANOVA, can be obtained using the entries from Table 17.2 similarly to the way they are obtained

in Chapter 16 for Table 16.3. The resulting ANOVA is presented in Excel's ANOVA in Table 17.5 of Section 17.6.

17.6 Excel calculations for the ANOVA table: No replication

The numerical entries for Table 17.3 can be made using Excel. Under "Tools" (in some versions of Excel under "Data") select "Data Analysis," and then select "ANOVA: Two-Factor Without Replication." In the ANOVA panel, select the data array. If the data contain labels as column headers, check the box "Labels in First Row." Then enter the level of significance (defaults to 0.05), and indicate the spreadsheet cell where the printout should start.

Two tables show Excel's output for the data in Example 17.1. Table 17.5 displays selected summary (descriptive) statistics, while Table 17.6 displays the ANOVA table.

Table 17.4. Excel's summary statistics for Example 17.1

<i>Summary</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
First	2	104.0	52.0	8.0
Middle	2	48.0	24.0	8.0
Last	2	49.0	24.5	0.5
Column 1	3	96.0	32.0	244.0
Column 2	3	105.0	35.0	271.0

The Excel printout continues with the ANOVA (called "Single Factor" by Excel).

Table 17.5. Excel's ANOVA table for Example 17.1

<i>Source of variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	1027.0	2	513.5	342.3	0.003	19.00
Columns	13.5	1	13.5	9.0	0.095	18.51
Error	3.0	2	1.5			
Total	1043.5	5				

Because this is a fixed effects design, the hypotheses are given by Equations (17.6) and (17.8):

Similarly to its printout in Chapter 16, Excel gives the critical point for the selected α (here, $\alpha = 0.05$) in the column labeled “ F_{crit} ,” thus eliminating the need for table lookup and interpolation. In addition, Excel gives the “ P -value,” which is the probability of obtaining an F -value at least as large as the one that was actually observed, assuming that the null hypothesis is true. A P -value is an indication of the strength of the test result. If $P < \alpha$, the null hypothesis is rejected.

In Example 7.1, the calculated F statistic for “Rows” is 342.3. The critical value for testing equality of row means (Equation 17.5) is $f_{0.95}(2,2) = 19.00$. Because $342.2 > 19.00$, the null hypothesis of row equality is rejected and we conclude that there is a row effect and that proximity to the reactor affect the sublimation rate. The calculated F statistic for “Columns” is 9.0. The critical value for testing equality of column means (Equation 17.7) is $f_{0.95}(1,2) = 18.51$. Because $9.0 < 18.51$, the null hypothesis of column equality is not rejected, citing no statistical evidence to claim that ice sublimation varies among the selected (fixed) columns.

We reemphasize that since both rows and columns were selected by design the study conclusions apply only to the selected columns and rows. However, had we selected the columns at random, the ANOVA calculations would have been no different. However, the conclusion would have been that we do not reject the null hypothesis that $\sigma_B^2 = 0$.

17.7 Balanced two-way factorial designs with replication

Many experiments are designed with more than one observation per experimental unit. In this section, we consider balanced designs where the number of replications is the same for every experimental unit. We denote the number of replications by n , where $n > 1$.

In a two-way factorial design with replication, we still have two main factors, A and B , with a and b levels, respectively. With n replications, the number of observations is $n_+ = abn$. By incorporating replication into the design, we can introduce another concept, namely interaction, into the model. Interaction is a measure of the simultaneous effect of the main effects, A and B .

In Example 17.2, we illustrate the concept of interaction with a modification of Example 17.1.

Example 17.2. Ice sublimation with interaction. The context for this example is the same as that of Example 17.1. As in Example 17.1, we weigh ice baskets in three rows and two columns. Suppose first that Figure 17.1 shows sublimation, as measured by the mean weight loss. Here, Row 1 is closest to and Row 3 is farthest from the reactor. Clearly, sublimation decreases with distance from the reactor. The pattern is the same for both columns, with Column 2 sublimation 20 kg larger than Column 1 sublimation for each row. Because the differences between the rows (levels of Factor *A*) are the same for each column (levels of Factor *B*), there is no interaction between the rows and the columns (i.e., between the factors).

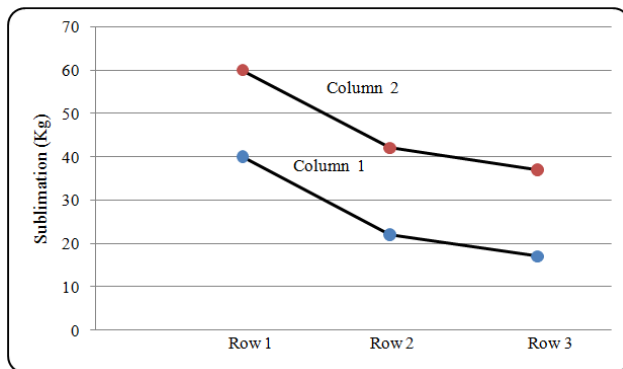


Figure 17.1. Sublimation in the absence of interaction

Suppose now that Figure 17.2 also gives the sublimation data. Comparison with Figure 17.1 shows that the results for Rows 1 and 2 remain unchanged, but are interchanged for Row 3, where the Column 2 sublimation is now 20 kg smaller than the Column 1 sublimation. We now have interaction, because the differences between the rows (levels of Factor *A*) depend on the columns (levels of Factor *B*).

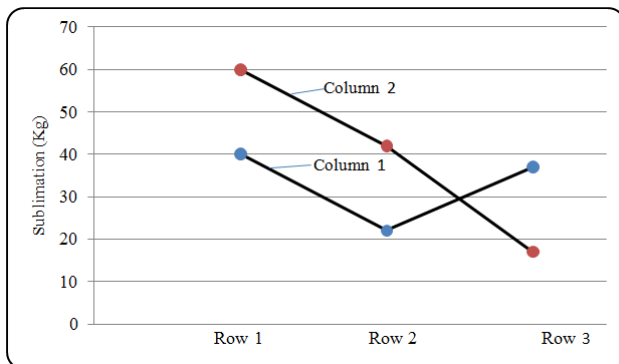


Figure 17.2. Sublimation with interaction

Let Y_{ijk} be the k^{th} observation at the i^{th} level of Factor A and the j^{th} level of Factor B in cell (i, j) . Extending Equation (17.1), the model for a two-way balanced factorial design with replication is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n \quad (17.13)$$

where:

μ = overall mean response

α_i = main effect of the i^{th} level of Factor A averaged over the levels of Factor B

β_j = main effect of the j^{th} level of Factor B averaged over the levels of Factor A

$(\alpha\beta)_{ij}$ = effect of the interaction between the i^{th} level of Factor A and the j^{th} level of Factor B

ε_{ijk} = random error associated with Y_{ijk}

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

- Some writers use a different symbol (e.g., γ) to denote two-factor interaction. We treat $(\alpha\beta)$ as if it were a single Greek letter to indicate that it refers to the joint effect of two factors.

The null and alternative hypotheses for this design are the same as those stated in Equations (17.5) through (17.11) with the addition of hypotheses for interaction. For a fixed effects model, these are:

$$H_0: (\alpha\beta)_{ij} = 0 \text{ all } i, j \text{ vs. } H_1: (\alpha\beta)_{ij} \neq 0, \text{ some } i, j \quad (17.16)$$

With replications, the sum of squares for total variability is partitioned as follows:

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error} \quad (17.17)$$

where AB denotes the interaction between Factors A and B .

Equations (17.18) gives the degrees of freedom associated with each of the sums of squares.

$$\begin{aligned}
 v_{Total} &= n_+ - 1 \\
 v_A &= a - 1 \\
 v_B &= b - 1 \\
 v_{AB} &= (a - 1)(b - 1) \\
 v_{Error} &= ab(n - 1)
 \end{aligned}
 \tag{17.18}$$

Table 17.6 shows the ANOVA table for a two-way design with replication.

Table 17.6. ANOVA table for a two-way factorial design with replication

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Factor A	SS_A	$v_A = a - 1$	$MS_A = \frac{SS_A}{v_A}$	F_A
Factor B	SS_B	$v_B = b - 1$	$MS_B = \frac{MS_B}{v_B}$	F_B
Interaction AB	SS_{AB}	$v_{AB} = (a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{v_{AB}}$	$F_{AB} = \frac{MS_{AB}}{MS_{Error}}$
Error	SS_{Error}	$v_{Error} = ab(n-1)$	$MS_{Error} = \frac{MS_{Error}}{v_{Error}}$	
Total	SS_{Total}	$v_{Total} = abn - 1$		

The test statistic for interaction effect, regardless of which main effect is fixed and which is random, is:

$$F_{AB} = MS_{AB}/MS_{Error} \tag{17.19}$$

The F -statistics for the main effects, Factor A and Factor B , depend on whether the effects are fixed, random, or mixed. When both Factors A and B are fixed effects, the F -statistics are:

$$F_A = MS_A/MS_{Error} \text{ and } F_B = MS_B/MS_{Error} \tag{17.20}$$

When both Factors A and B are random effects, the F -statistics are:

$$F_A = MS_A/MS_{AB} \text{ and } F_B = MS_B/MS_{AB} \tag{17.21}$$

When Factor A is a fixed effect and Factor B is a random effect, the F -statistics are:

$$F_A = MS_A/MS_{AB} \text{ and } F_B = MS_B/MS_{Error} \tag{17.22}$$

Example 17.3. Lathe and coolant interaction (“NIST Engineering Statistics Handbook” (2006), Section 3.2.3.2.1). A two-way factorial experiment is designed to test for differences in turned pin diameters (in inches) due to the use of different lathes or different coolants. In this experiment, five samples are taken from each of five lathes, each using two different types of coolant. Table 17.7 gives the data in the format required by Excel. Because there are only five lathes and two coolants, this is a fixed effects design.

Table 17.7. Two-factor data for Example 17.3 in Excel format

	A	B	C	D	E	F
1		Lathe 1	Lathe 2	Lathe 3	Lathe 4	Lathe 5
2	Coolant A	0.125	0.118	0.123	0.126	0.118
3	Coolant A	0.127	0.122	0.125	0.128	0.129
4	Coolant A	0.125	0.120	0.125	0.126	0.127
5	Coolant A	0.126	0.124	0.124	0.127	0.120
6	Coolant A	0.128	0.119	0.126	0.129	0.121
7	Coolant B	0.124	0.116	0.122	0.126	0.125
8	Coolant B	0.128	0.125	0.121	0.129	0.123
9	Coolant B	0.127	0.119	0.124	0.125	0.114
10	Coolant B	0.126	0.125	0.126	0.130	0.124
11	Coolant B	0.129	0.120	0.125	0.124	0.117

The two factors involved are lathe, with five levels, and coolant, with two levels. Each lathe \times coolant combination has five replications. Excel’s analysis requires that the number of replications be the same for each such combination (i.e., that the design be balanced).

Excel requires that all data entry headers must be included in the data range. In Table 17.7, column headers are shown in cells B1..F1. Row headers are shown in cells A2..A11, although they do not have to be repeated for every replication. Thus, we could enter “Coolant A” and “Coolant B” in cells A2 and A7 and leave the rest of the cells in Column A empty.

Call Excel’s ANOVA routine as follows. Under “Data” (in some versions of Excel, under “Tools”), select “Data Analysis,” and then select “ANOVA: Two-Factor With Replication.” In the ANOVA panel, select the data array, which must include column and row headers (here, A1, ..., F11). Under “Rows per sample,” enter the number of replications ($n = 5$, in this example). Select the level of significance (defaults to $\alpha = 0.05$).

Two tables show Excel’s output for the data from Table 17.7. First, Table 17.8 displays some summary (descriptive) statistics.

Table 17.8. Excel's summary statistics for Example 17.3

SUMMARY	Lathe 1	Lathe 2	Lathe 3	Lathe 4	Lathe 5	Total
<i>Coolant A</i>						
Count	5	5	5	5	5	25
Sum	0.6310	0.6030	0.6230	0.6360	0.6150	3.1080
Average	0.1262	0.1206	0.1246	0.1272	0.1230	0.1243
Variance	1.70E-06	5.80E-06	1.30E-06	1.70E-06	2.25E-05	1.12E-05
<i>Coolant B</i>						
Count	5	5	5	5	5	25
Sum	0.6340	0.6050	0.6180	0.6340	0.6030	3.0940
Average	0.1268	0.1210	0.1236	0.1268	0.1206	0.1238
Variance	3.70E-06	1.55E-05	4.30E-06	6.70E-06	2.33E-05	1.64E-05
<i>Total</i>						
Count	10	10	10	10	10	
Sum	1.2650	1.2080	1.2410	1.2700	1.2180	
Average	0.1265	0.1208	0.1241	0.1270	0.1218	
Variance	2.50E-06	9.51E-06	2.77E-06	3.78E-06	2.20E-05	

Table 17.9 shows the Excel printout of the ANOVA. Note that Excel's ANOVA also provides the critical values for the analysis, eliminating the need for table lookup and interpolation.

Table 17.9. Excel's ANOVA for Example 17.3

Source of Variation	SS	Df	MS	F	P-value	F crit
Coolant	3.92E-06	1	3.92E-06	0.45	0.50470	4.08
Lathes	0.000303	4	7.58E-05	8.77	0.00004	2.61
Interaction	1.47E-05	4	3.67E-06	0.42	0.79017	2.61
Error	0.000346	40	8.65E-06			

Because $0.45 < 4.08$, the coolant effect is not significant. Because $8.77 > 2.61$, the lathe effect is statistically significant. Because $0.42 < 2.61$, the coolant \times lathe interaction is not statistically significant.

17.8 Other multifactor designs

Many other experimental designs are used for ANOVA, such as nested (hierarchical), fractional factorial, and Latin square designs, both balanced and unbalanced. These topics are beyond the scope of this book. Interested readers can find details in Bowen and Bennett (1988) or Neter et al. (1990).

18

Regression

18.1 What to look for in Chapter 18

This chapter introduces the topic of regression, a set of statistical procedures designed to investigate the relationships among two or more variables. The first part is restricted to one dependent and one independent variable. In Section 18.16, we consider one dependent variable and several independent variables.

In developing this subject, we will revisit some terms from algebra and geometry:

- *independent and dependent variables, §18.2*
- *slope and intercept, §18.2*

We will also revisit some concepts and terms from the analysis of variance (Chapters 16 and 17) that play a similar role in regression analysis:

- *model and error, §18.4*
- *sum of squares and mean square error, §18.7*
- *partition of sum of squares, §18.9*
- *regression analysis of variance, §18.10*

We also introduce the following concepts and ideas:

- *curve fitting*, §18.3
- *simple linear regression*, §18.5
- *least squares*, §18.7
- *testing the slope and confidence interval for the slope*, §18.13
- *testing the intercept and confidence interval for the intercept*, §18.14
- *regression through the origin*, §18.15
- *multiple linear regression*, §18.16
- *prediction and prediction interval*, §18.17

We also show detailed steps for using Excel:

- *for plotting a regression line*, §18.11
- *for regression analysis*, §18.12

18.2 Concepts and terms from algebra and geometry

independent variable At some point in our mathematics training, we learned to plot mathematical functions of various types. We usually started with an ***independent variable***, usually called x , which was related to a ***dependent variable***, usually called y . The dependence of y on x was expressed as the function $y = f(x)$. We learned to build a table of ***ordered pairs*** of values, denoted by (x, y) , and then to plot those pairs and connect them on a coordinate system.

To illustrate, suppose we are asked to plot the function $y = 3x + 2$. We pick some convenient values for the independent variable x and calculate the corresponding values of the dependent variable y . Table 18.1 displays some selected (x, y) pairs for this function.

Table 18.1. Ordered pairs (x, y) for the function $y = 3x + 2$

x	-2	0	1	2	5
y	-4	2	5	8	17

The (x, y) pairs are used to produce a graph of the given function, as shown in Figure 18.1.

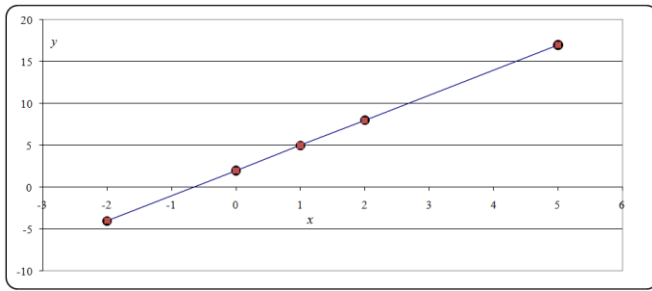


Figure 18.1. Graph of the function $y = 3x + 2$

The general equation of a straight line in an x, y coordinate system is $y = mx + b$, where m is the *slope* and b is the *intercept* (sometimes called the y -intercept). Thus, for the line $3x + 2$, the slope is 3, meaning that y increases by 3 units for every unit increase in x . For the same line, the intercept is 2, which is the value of y for which $x = 0$. Notice especially that the line passes through each of the points corresponding to the ordered pairs.

- The coefficients m and b in the function $y = mx + b$ are the parameters of the general straight line. In a nonstatistical context, these parameters are written as lowercase Latin letters. In contrast, we write statistical parameters as lowercase Greek letters, estimators of statistical parameters as uppercase Latin letters, and estimates of statistical parameters as lowercase Latin letters.

Because two points determine a straight line, why did we use five points for plotting the line? Operationally, it is simply good practice to plot a few extra points to be certain that we set up the graph correctly.

18.3 The concept of regression

Example 18.1. Five points in search of a trend. Consider Table 18.2, the points of which are plotted in Figure 18.2. Although there is no straight line that passes through all of these points, it is often useful to find a straight line that fits the points reasonably well. This is the problem that regression is designed to solve.

Table 18.2. Five ordered pairs (x, y)

x	1	2	3	5	7
y	4	3	4	8	9

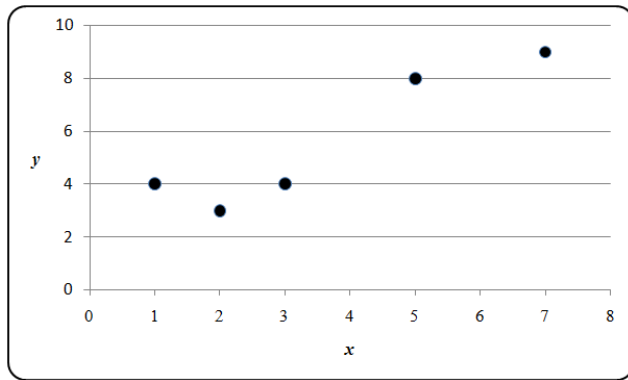


Figure 18.2. Graph of the data in Table 18.2

The difference between Figures 18.1 and 18.2 reflects the essence of the regression approach. In an algebraic or geometric setting, we know the function that links the values of y to the values of x , and our analysis is based on that knowledge. By contrast, in a statistical setting, we are given the (x, y) pairs, and our task is to find a mathematical function that links the ordered pairs. For example, the values of y in Figure 18.2 tend to increase as x increases, and regression can be used to find a straight line that expresses this trend.

regression The process of determining an approximate algebraic relationship between two or more variables is called **regression**.⁸ Because **curve fitting** regression fits a model (an equation, a line, or a curve) to a set of points, regression is sometimes called **curve fitting**.

At the risk of stating the obvious, statistical concepts and tools cannot perform miracles. We cannot produce a straight line that goes through every one of an arbitrary set of points if the points are not already on a straight line. Rather, regression produces a compromise line that meets some reasonable criteria.

⁸ According to Kendall and Buckland (1971), p. 127, the term regression "...was originally used by Galton [Sir Francis Galton, 1822–1911, an English biostatistician and cousin of Charles Darwin] to indicate certain relationships in the theory of heredity but it has come to mean the statistical method developed to investigate those relationships."

18.4 Regression models

statistical model Regression begins with a *statistical model* that is a symbolic representation of the relationship between the variables involved in a study. The model incorporates a functional relationship affected by unexplained, unknown, or random fluctuations. In this chapter, we focus on a statistical model that relates a dependent variable, denoted by an uppercase Y , to an independent variable, designated by a lowercase x . We assume that the independent variable is measured without, or with negligible, error.

As an example, consider the relation between time and distance as we drive on a road. Whenever we pass a mile marker we record the distance covered since the beginning of the trip and the time it took to travel that distance. The distance traveled is known, but the time depends on factors such as the weather and traffic conditions and is therefore a random variable. The distance, being fixed, is denoted by x and the time, being a random variable, is denoted by Y . Here, x is the independent variable and Y is the dependent variable. We wish to find a linear function of the form $Y = \alpha + \beta x$ that relates the dependent to the independent variable. In many applications, this function is used to predict the value of Y corresponding to a given value of x . The equation $Y = \alpha + \beta x$ is, at best, an imperfect representation of data that do not all lie on a straight line, such as those in Table 18.2. To account for this imperfection, we add an

error *error* term to the model. Denoting the error by ε_i , we have the model:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (18.1)$$

where:

Y_i = the value of the dependent variable for the i^{th} point

x_i = the value of the independent variable for the i^{th} point

α = an unknown constant (intercept)

β = an unknown constant (slope)

n = number of (x_i, Y_i) points

ε_i = the error associated with the i^{th} point

The error ε_i reflects the failure of the i^{th} point to lie on the line described by the model. This error may be attributed to experimental error, experimenter error, modeling error, random error, or a combination of these and other errors.

A model with a single independent variable, such as in Equation (18.1) is said to be *simple*. If, in addition, the model is additive, then the model is also said to be *linear*. The procedure for finding a line using such a model is called ***simple linear regression***. Simple linear regression is the main topic of this chapter.

A more complex model that involves several independent variables such as:

$$Y_i = \alpha + \beta x_i + \gamma w_i + \delta z_i + \dots + \varepsilon_i \quad (18.2)$$

multiple linear regression is analyzed by a ***multiple linear regression*** procedure, which is addressed in Section 18.16.

If the independent variable in the model is raised to a power of 2 or higher, such as:

$$Y_i = \alpha + \beta x_i + \gamma x_i^2 + \delta x_i^3 + \dots + \varepsilon_i \quad (18.3)$$

polynomial regression the associated regression procedure is called ***polynomial regression***.

If the logarithm of the dependent variable is a linear function of the independent variable, such as:

$$\log(Y_i) = \alpha + \beta x_i + \dots + \varepsilon_i \quad (18.4)$$

exponential regression the associated regression procedure is called ***exponential regression***.

Other complex models are found in the literature. For detailed discussion of such models, see Neter, et al. (1996), Chapters 6 and 7.

Some writers use a set of coefficients $\{\beta_k\}$ with different subscripts instead of α , β , γ , \dots . Thus, the simple linear regression in Equation (18.1) uses β_0 for the intercept and β_1 for the slope, as shown below:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (18.5)$$

Two advantages of the model in Equation (18.5) over that in Equation (18.1) are usually cited:

- There is no confusion of the coefficients, α and β , with the probabilities of Type I and Type II errors in hypothesis testing.
- The model can be extended to include many variables and combinations, such as $Y = \beta_0 + \beta_{1x} + \beta_{2w} + \beta_{3z} + \beta_{12xw} + \beta_{xx}^2 + \dots + \varepsilon_i$ without exhausting the Greek alphabet.

In dealing with simple linear regression, we prefer to use the model in Equation (18.1) because it is more commonly used when a single independent variable is considered.

Note that the constant α in Equation (18.1) may be zero, in which case the model becomes:

$$Y_i = \beta_{xi} + \varepsilon_i, i = 1, 2, \dots, n \quad (18.6)$$

regression through the origin This last model describes a line that goes through the origin (i.e., a zero-intercept line). The process for constructing such a line is referred to as **regression through the origin**. Many studies require such a model (e.g., the time and distance example at the beginning of this section). Clearly, because it takes 0 time to go 0 distance, the regression line should go through the origin.

18.5 Simple linear regression

empirical line Given data that we assume follow the model in Equation (18.1), we estimate the model parameters (i.e., the intercept and slope) and construct an **empirical line** that captures the relation between the independent and dependent variables. If we use a regression process to construct an empirical line, it is called a **regression line**.
regression line For a simple linear regression, the regression line is written as:

$$\hat{Y}_i = A + Bx_i, i = 1, \dots, n \quad (18.7)$$

where:

A (sometimes written as $(\hat{\alpha})$) is an estimator of the model intercept α .

B (sometimes written as $(\hat{\beta})$) is an estimator of the model slope β .

prediction equation The line in Equation (18.7) is also called the **prediction equation** because it can be used to predict a value of Y for a fixed value of x .

- The term \hat{Y}_i , pronounced *why-hat-sub-eye*, is an estimator of the true-but-unobserved random variable Y_i and thus carries a “hat” (caret) to distinguish it from the observed y_i . This notation follows the widely used practice of “hatting” a symbol to distinguish an estimate from the value we are trying to estimate.

The estimators of the intercept and the slope, A and B , are functions of the sample we happen to collect and of our method of estimation. Because the sample contains random variables, it follows that A and B are also random variables.

Any statistical treatment of data is based on a collection of assumptions about the dataset and its structure. The assumptions for simple linear regression are of two types:

- basic assumptions for the construction of the regression coefficients
- one additional assumption for tests of hypotheses and confidence intervals

The additional assumption is not needed for construction of the regression line, but it is needed for statistical analysis of the regression line.

The basic assumptions for simple linear regression are:

- The data follow the model in Equation (18.1).
- Each x_i is measured without error, or with negligible error.
- The variance of Y_i is a constant, denoted by σ^2 , for all x_i values. This means that the variance of the population of all possible Y observations for any value of x is σ^2 .
- If the latter assumption cannot be met, then construction of a regression line requires weighted regression techniques that are beyond the scope of this book. See Draper and Smith (1981) or Neter, et al. (1996) for an extensive treatment of this subject.

The additional assumption needed for testing hypotheses or constructing confidence intervals for the model parameters is that the error ε_i associated with Y_i is normally distributed with mean 0 and variance σ^2 . Symbolically, we write:

$$\varepsilon_i \sim N(0, \sigma^2) \quad (18.8)$$

Figure 18.3 illustrates these assumptions.

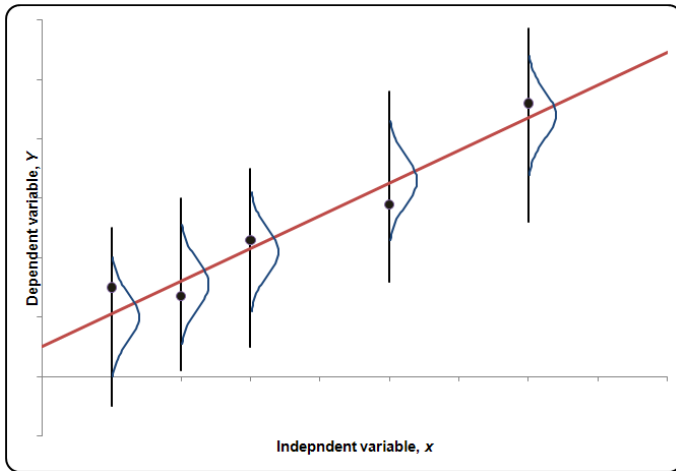


Figure 18.3. An Illustration of the assumptions for simple linear regression

From Figure 18.3, we see the following:

- Each of the five data points in the figure may be regarded as a bead capable of sliding on a fixed string (vertical line). A bead cannot slide in a horizontal direction, reflecting the assumption that each x is measured without error.
- In contrast with the no-error assumption about the x values, the Y 's are random variables. That is, the beads can slide vertically, following their distribution.
- A normal distribution with a constant variance is associated with every x value.

Finally, the straight line superimposed on the figure illustrates the fitted regression line for the data points (see Section 18.7).

18.6 Fitting a line to data

In the following sections, we provide the basis for the construction of regression lines and identify the components of the regression analysis. Readers who do not need this background may turn to Sections 18.11 and 18.12, which discuss the use of Excel to perform regression graphics and analysis.

The model is given by Equation (18.1). The intercept and the slope, α and β , are unknown constants to be estimated from the data. Given a set of n (x_i, y_i) points, there are many ways to find a line that fits these points. For instance, we could plot the data and simply draw a line that seems to fit (i.e., do an “eyeball” fit). While such lines are usually pretty good fits, this procedure is unacceptable

because it is not reproducible. Eyeball fit is very likely to yield different lines if done by different analysts or even by the same analyst at different times. Consequently, it would not be amenable to statistical analysis.

Clearly, an objective and reproducible procedure is needed. One possibility is to use a horizontal line that goes through the mean \bar{y} of the observations. In particular, if the dependent and independent variables are unrelated, this line would make sense. For example, if the dependent variable Y represents the yearly energy consumption by a household and the independent variable x represents the cost of tea, it makes sense to ignore the cost of tea and estimate the energy consumption of the i^{th} household by $\hat{Y}_i = \bar{y}$ (see Equation (18.7)). Moreover, even when x and Y are considered related for theoretical or practical reasons, it often is useful to use the horizontal line $\hat{Y}_i = \bar{y}$ as a baseline, or standard, against which other lines may be compared.

We now return to the five points in Table 18.2. Their scatter diagram is plotted in Figure 18.4, which is the same as Figure 18.2 but with a horizontal line superimposed whose equation is $Y_i = \bar{y} = 5.6$. This line is our first candidate to fit the data in Table 18.2.

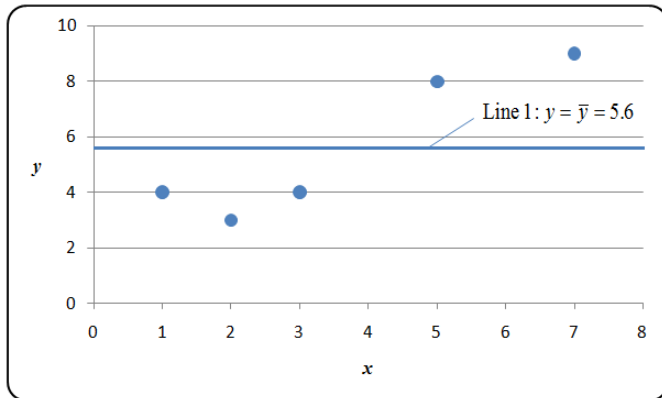


Figure 18.4. Five points from Table 18.2 and the line $y = \bar{y}$

Figure 18.5 adds two more lines for consideration. Line 2 is the equation $y = -2 + 2x$, and Line 3 is the equation $y = 2 + (8/9)x$.

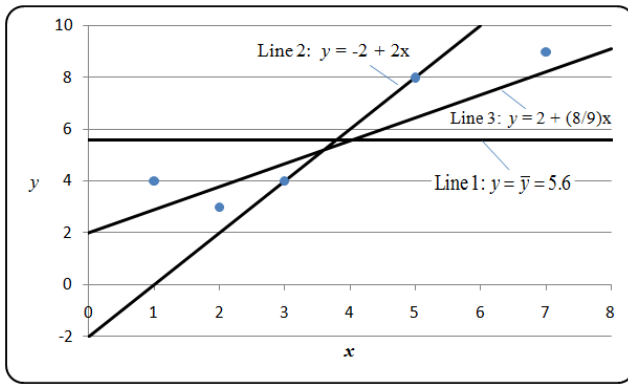


Figure 18.5. Three candidate lines for fitting the data in Table 18.2

Which of the three candidate lines, if any, should be picked? Some candidates are “better” fits than others, and some are “worse” fits. Clearly, we need a criterion for choosing the “best” line, and this is the subject of the next section.

18.7 The method of least squares and the regression line

The criterion we shall use for the selection of a best line is based on a definition of the “distance” from the data points to any candidate line for fitting the data. We illustrate the distance from the data points to a line with the three lines in Figure 18.5. Consider Line 3. For each of the five points, draw a vertical line that connects the point to Line 3, as shown in Figure 18.6. The length of each line is the vertical distance between the point and Line 3.

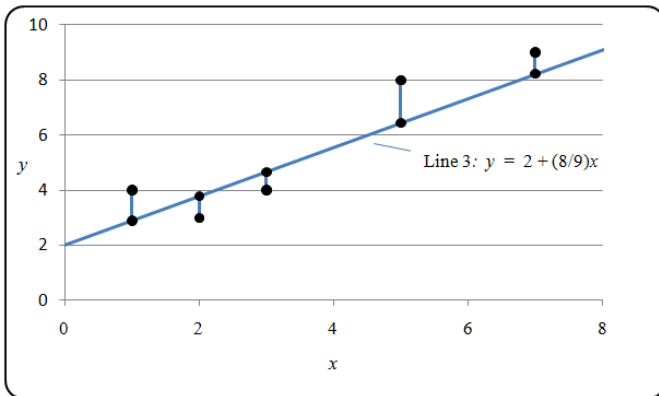


Figure 18.6. Vertical distances from Line 3

Next, square each distance, add the squares to obtain a sum of squares of vertical deviations of the points from Line 3, and denote it by SS_3 . This sum of squares

is the distance between Line 3 and the data points. Repeat the process for Lines 1 and 2 to form SS_1 and SS_2 , respectively. Table 18.3 shows the results of these calculations.

Table 18.3. Sums of squared deviations for the three lines in Figure 18.5

Data						Sum of Squares
X	1	2	3	5	7	
Y	4	3	4	8	9	
Line 1: $y = 5.6$						$SS_1 = 29.2000$
\hat{y}_i	5.60	5.60	5.60	5.60	5.60	
$y_i - \hat{y}_i$	-1.60	-2.60	-1.60	2.40	3.40	
$(y_i - \hat{y}_i)^2$	2.5600	6.7600	2.5600	5.7600	11.5600	
Line 2: $y = 2x - 2$						$SS_2 = 26.0000$
\hat{y}_i	0.00	2.00	4.00	8.00	12.00	
$y_i - \hat{y}_i$	4.00	1.00	0.00	0.00	-3.00	
$(y_i - \hat{y}_i)^2$	16.0000	1.0000	0.0000	0.0000	9.0000	
Line 3: $y = (8/9)x + 2$						$SS_3 = 5.3086$
	2.8889	3.7778	4.6667	6.4444	8.2222	
$y_i - \hat{y}_i$	1.1111	-0.7778	-0.6667	1.5556	0.7778	
$(y_i - \hat{y}_i)^2$	1.2346	0.6049	0.4444	2.4198	0.6049	

In general, the distance D between a line and a set of points is defined as the sum of the squared vertical deviations between the line and the points. Using this definition, Line L_1 is a better fit to a dataset than Line L_2 if D_1 is smaller than D_2 . Of the three lines in Figure 18.5, Line 3 is the best fit because it has the smallest sum of squares ($SS_3 = 5.3086$). However, there might be another line that has a smaller distance to the data and would therefore be a better fit to the data. Clearly, our goal is to find a best line to fit the data (i.e., a line that minimizes the distance to the data points).

least squares line Using regression methodology, we can readily construct such a line. For any dataset with more than one point, there is only one such line, and it is called the **least squares line**. The procedure for constructing a least squares line is the **method of least squares**.

method of least squares Details of this method may be found in Neter, et al. (1996), Chapter 1; Draper and Smith (1981), Chapter 2; and many other sources.

The least squares approach to fitting a line to data is only one of several procedures called regression. However, the least squares approach is, by far, the most common form of regression. Throughout this book, we use the term regression to mean a

regression line procedure for fitting a line (or sometimes a curve) to a dataset by the method of least squares. This line is called the least squares line or the **regression line**.

The method of least squares, as applied to a simple linear model, can be described as follows.

We have a set of n ordered pairs $\{(x_i, Y_i)\}$, $i = 1, 2, \dots, n$, where $n > 1$ and at least two of the x values are different. Then there is a unique least squares line with parameters given by Equation (18.7). The formula for the estimator B of the slope is given by two mathematically equivalent equations. The first is the “definition formula,” given by Equation (18.9), and the second is the “working formula,” given by Equation (18.10). The working formula is almost always more amenable to statistical calculations. It is the formula used in hand-held calculators and in most statistical packages.

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (18.9)$$

$$B = \frac{\sum_{i=1}^n x_i Y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n Y_i) / n}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 / n} \quad (18.10)$$

Once the slope is calculated, the estimator A of the slope is obtained as:

$$A = \bar{Y} - B\bar{x} \quad (18.11)$$

- Mathematically, it can be shown that the regression line goes through the point (\bar{x}, \bar{y}) . This bit of information could be useful to hand-draw the regression line through a dataset.

We introduce the following notation; where each summation runs from $i=1$ to $i=n$

$$S_{xx} = \sum (x - \bar{x})^2 = \sum x^2 - \sum x^2 / n = \sum x^2 - n\bar{x}^2 \quad (18.12)$$

$$S_{xY} = \sum (x - \bar{x})(Y - \bar{Y}) = \sum xY - \sum x \sum Y / n = \sum xY - n\bar{x}\bar{Y} \quad (18.13)$$

$$S_{YY} = \sum (Y - \bar{Y})^2 = \sum Y^2 - \sum Y^2 / n = \sum Y^2 - n\bar{Y}^2 \quad (18.14)$$

Using this notation, the formula for the slope is written as:

$$B = S_{xy}/S_{xx} \quad (18.15)$$

Note the mixture of subscripts in S_{xy} , S_{xY} , and S_{YY} in Equations (18.12) through (18.15).

As these formulas are applied to the data, the random variables (written in uppercase) are replaced by the numerical values of their observations (written in lowercase).

To construct the least squares line for Example 18.1, we start with some preliminary calculations:

$$\Sigma x = 1 + 2 + 3 + 5 + 7 = 18$$

$$\Sigma x^2 = 1^2 + 2^2 + 3^2 + 5^2 + 7^2 = 88$$

$$\Sigma y = 4 + 3 + 4 + 8 + 9 = 28$$

$$\Sigma y^2 = 4^2 + 3^2 + 4^2 + 8^2 + 9^2 = 186$$

$$\Sigma xy = (1)(4) + (2)(3) + (3)(4) + (5)(8) + (7)(9) = 125$$

$$S_{xx} = 88 - (18)^2/5 = 23.20$$

$$S_{yy} = 186 - (28)^2/5 = 29.20$$

$$S_{xy} = 125 - (18)(28)/5 = 24.20$$

- Although the term S_{YY} is not used in the calculation of A or B , S_{YY} comes into play at later stages, when the statistical properties of the regression line are examined and when correlation methods are used (Chapter 19).

From Equations (18.15) and (18.11), the regression coefficients are:

$$B = 24.20/23.20 = 1.0431$$

$$A = 28/5 - (1.0431)(18/5) = 1.8448$$

Thus, the regression line for Example 18.1 is:

$$y = 1.8448 + 1.0431x$$

The calculations in Table 18.4 yield the sum of squares = 3.9569. This sum of squares is smaller than any sum of squares in Table 18.3, and because it is the sum of squares for the least squares line, we are guaranteed that we cannot find another set of A and B that would give us a smaller sum of squares.

Table 18.4. The least squares line and sum of squares for Example 18.1

Data						
<i>x</i>	1	2	3	5	7	Sum of Squares
<i>y</i>	4	3	4	8	9	
Least squares line: $y = 1.8448 + 1.0431x$						
\hat{y}_i	2.8879	3.9310	4.9741	7.0603	9.1465	SS = 3.9569
$y_i - \hat{y}_i$	1.1121	-0.9310	-0.9741	0.9397	-0.1465	
$(y_i - \hat{y}_i)^2$	1.2368	0.8668	0.9489	0.8830	0.0215	

18.8 Geometric interpretation of regression components

The following identity holds for the difference between any observation and the sample mean:

$$(Y_i - \bar{Y}) \equiv (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \tag{18.16}$$

total deviation The term $(Y_i - \bar{Y})$ represents the **total deviation** of the i^{th} observation y_i from the mean \bar{Y} of all the observations. From Equation (18.16) it is the sum of two components:

- deviation due to regression** ▪ $(\hat{Y}_i - \bar{Y})$ that represents **deviation due to regression**
- deviation due to error** ▪ $(Y_i - \hat{Y}_i)$ that represents the **deviation due to error**

In terms of the deviations, Equation (18.16) can be written as:

$$\text{total deviation} = \text{deviation due to regression} + \text{deviation due to error} \tag{18.17}$$

residual The deviation due to error is often called the **residual**. Intuitively, the residuals are the parts of the observations that the regression line does not explain.

Equation (18.16) has a geometric interpretation. This interpretation is illustrated in Figure 18.7 at the fourth observation from Table 18.2, where $(x_4, y_4) = (5, 8)$. From Table 18.4, the fitted value \hat{y}_4 is 7.06, yielding the fitted point (5, 7.06).

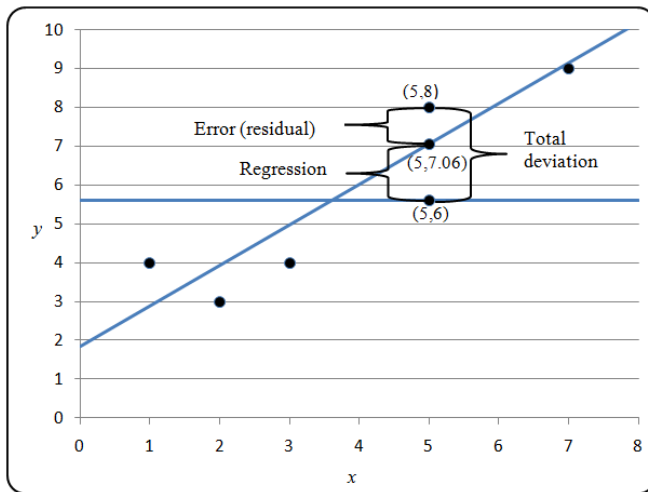


Figure 18.7. Geometric interpretation of regression components

The total deviation ($y_4 - \bar{y}$) is the vertical distance between the observed point $(x_4, y_4) = (5, 8)$ and the observation mean $\bar{y} = 5.60$. This deviation is calculated as $8 - 5.60 = 2.40$.

The deviation due to regression ($\hat{y}_4 - \bar{y}$) is the vertical distance between the fitted point $(5, 7.06)$ and the observation mean. This deviation is calculated as $7.06 - 5.60 = 1.46$. A small value for this deviation means that the regression line does little to improve the fit beyond that [which is] offered by the horizontal line $y = \bar{y}$, at least near this data point. A large value for this deviation means that the regression line improves the fit to the data point over that [which is] provided by the sample mean.

The deviation due to error is the residual $y_4 - \hat{y}_4$, which is the distance between the observed point $(5, 8)$ and the fitted point $(x_4, \hat{y}_4) = (5, 7.06)$. This deviation is calculated as $8 - 7.06 = 0.94$. A small value for a residual means that the regression line provides a reasonable fit to the observation. A large value for a residual means that the regression line does not fit the data well, at least in the neighborhood of the observed point.

18.9 Partition of the total sum of squares

A sensible question we may ask at this point is whether the effort involved in fitting a line to a given dataset is warranted. Putting the question another way: Is there evidence of a trend in the data? For a straight line, a trend is characterized by a non-zero slope, meaning that as x changes, there is a predictable change in Y . If there is no trend, then we can simply use the sample mean to predict Y regardless of the value of x .

To answer this and other questions, we perform a statistical analysis of the regression. From this point on, we need the assumption of a normal distribution of error as in Equation (18.8), because this assumption is the basis for the regression analysis. This analysis is based on a partition of the sum of squares of the observations about the sample mean into two statistically independent components.

Recall the identity $(Y_i - \bar{Y}) \equiv (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$ in Equation (18.16). The following remarkably similar expression, also an identity, is a cornerstone of the regression analysis:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 \equiv \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (18.18)$$

- The proof of this identity may be found in many textbooks on regression (e.g., Neter, et al. (1996), p. 72). Also, because there is no ambiguity, we will henceforth omit the subscripts in the summations.

While Equation (18.16) partitions the variation about the sample mean of an individual observation into two components, Equation (18.18) partitions the joint variation about the sample mean of all the sample points into two components. This partition is quite similar to the partition described in Chapters 16 and 17 for the analysis of variance.

total sum of squares The sum of squares $\sum (Y - \bar{Y})^2$ is called the **total sum of squares**, denoted by SS_{Total} . We have:

$$SS_{Total} = S_{YY} \quad (18.19)$$

where S_{YY} is given by Equation (18.14). The degrees of freedom associated with SS_{Total} is:

$$v_{Total} = n - 1 \quad (18.20)$$

From Equation (18.18), SS_{Total} is the sum of two other sums of squares:

regression sum of squares $\sum (\hat{Y} - \bar{Y})^2$ is the **regression sum of squares**, denoted by SS_{Reg} . From Equations (18.12) and (18.13), we have:

$$SS_{Reg} = S_{xY}^2 / S_{xx} \quad (18.21)$$

The degrees of freedom associated with SS_{Reg} is:

$$v_{Reg} = 1 \quad (18.22)$$

error sum of squares $\sum (Y - \hat{Y})^2$ is the *error sum of squares*, denoted by SS_{Error} . It is calculated as:

$$SS_{Error} = SS_{Total} - SS_{Reg} \quad (18.23)$$

The degrees of freedom associated with SS_{Error} is:

$$v_{Error} = n - 2 \quad (18.24)$$

residual sum of squares SS_{Error} is sometimes (Excel, for example) called the *residual sum of squares*.

Note that not only are the sums of squares additive, per Equation (18.18), but so are the degrees of freedom, that is:

$$v_{Total} = v_{Reg} + v_{Error} \quad (18.25)$$

18.10 Regression ANOVA for a single independent variable

The three different sources of variation and their corresponding sums of squares and degrees of freedom are now entered into a regression analysis table, as shown in Table 18.5.

Table 18.5. A regression ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Regression	SS_{Reg}	$v_{Reg} = 1$	$MS_{Reg} = \frac{SS_{Reg}}{v_{Reg}}$	$F = \frac{MS_{Reg}}{MS_{Error}}$
Error (residual)	SS_{Error}	$v_{Error} = n - 2$	$MS_{Error} = \frac{SS_{Error}}{v_{Error}}$	
Total	SS_{Total}	$v_{Total} = n - 1$		

Table 18.5 also includes the following entries:

regression mean square **Regression mean square**, $MS_{Reg} = SS_{Reg} / v_{Reg}$ (18.26)

error mean square **Error mean square**, $MS_{Error} = SS_{Err} / v_{Error}$ (18.27)

$$F = MS_{Reg} / MS_{Error} \quad (18.28)$$

mean square error The error mean square is often referred to as the **mean square error (MSE)**. This is the estimate of σ^2 in the model, Equation (18.8).

➤ Note that the ANOVA table does not display a mean square entry for the total sum of squares, as it is not needed.

The F -statistic in the last column of Table 18.5 is the test statistic for testing the slope of the regression line. The null and the alternative hypotheses are written as:

$$H_0: \beta = 0 \quad (18.29)$$

$$H_1: \beta \neq 0 \quad (18.30)$$

trend The alternative H_1 is equivalent to saying that there is a **trend** in the data, meaning that as x changes, there is a predictable change in Y .

If the calculated F -statistic is larger than the table value $F_{1-\alpha}(1, n-2)$, H_0 is rejected. Otherwise, H_0 is not rejected for lack of evidence to the contrary.

For Example 18.1, the sums of squares are calculated using the intermediate calculations from Section 18.7 and Equations (18.19), (18.21), and (18.23); and

$$SS_{Total} = S_{yy} = 29.20$$

$$SS_{Reg} = S_{xy}^2 / S_{xx} = (24.20)^2 / 23.20 = 25.24$$

$$SS_{Error} = SS_{Total} - SS_{Reg} = 29.20 - 25.24 = 3.96$$

Table 18.6 gives the regression ANOVA.

Table 18.6. Regression ANOVA table for Example 18.1

Source of variation	Sum of squares	Degrees of freedom	Mean square	<i>F</i>
Regression	25.24	1	25.24	19.14
Error (residual)	3.96	3	1.32	
Total	29.20	4		

From Table T-4 of the appendix, the critical value is $F_{0.95}(1,3) = 10.1$. Since the calculated $F = 19.14$ exceeds 10.1, we reject H_0 and conclude that $\beta \neq 0$ (i.e., there is a trend in the data).

18.11 Using Excel to construct the regression line

The following example illustrates how to use Excel's regression routines to plot and analyze data.

Example 18.2. Containment pressure stabilization. The pressure inside a pressurized containment was monitored for several hours to determine if the pressure changed significantly during that time. Three variables were recorded during the study, as shown in Table 18.7:

- time elapsed (hours) since the beginning of the study, recorded, including the labels, in cells A1..A12
- containment inside temperature (°F), recorded in cells B1..B12
- containment inside pressure (pounds per square inch (psi)), recorded in cells C1..C12

Table 18.7. Dataset for Example 18.2

	A	B	C
1	Time	Temp	Pressure
2	0	82.0	35.0
3	7	79.1	33.5
4	14	91.1	35.5
5	32	95.4	36.0
6	45	78.7	34.0
7	66	66.6	32.5
8	80	85.3	34.0
9	103	77.7	35.0
10	124	86.3	34.5
11	151	89.9	35.0
12	180	74.3	32.0

In this example, both time and temperature are independent variables, and pressure is the dependent variable. For ease of presentation, we divide the example into two parts:

- **Example 18.2a**, where the independent variable is time
- **Example 18.2b**, where the independent variable is temperature

Excel offers some simple functions for constructing a regression line. For the intercept and slope, use the following function calls:

- =INTERCEPT(Range of y values, Range of x values). For Example 18.2a, use Table 18.7 and =INTERCEPT(C2..C12, A2..A12) = 34.828. For Example 18.2b, use Table 18.7 and =INTERCEPT(C2..C12, B2..B12) = 24.078. Note that this function does not include variable labels in the data range.
- =SLOPE(Range of y values, Range of x values). For Example 18.2a, use Table 18.7 and =SLOPE(C2..C12, A2..A12) = -0.008. For Example 18.2b, use Table 18.7 and =SLOPE(C2..C12, B2..B12) = 0.124. Note that this function does not include variable labels in the data range.

To plot the data points in addition to constructing the regression line, use the Excel routine for regression graphics with the 10 steps below. Step 1 applies to Example 18.2a.

- Step 1.** Select the independent and dependent variables. Since time and pressure are not listed contiguously, we select them using the Ctrl key. Hold down the Ctrl key, and select cells A2..A12. Continue to hold down the Ctrl key, and select cells C2..C12. If the variable names are included in the row above the data (as in columns A1 and B1), we may select them too, as in A1..A12, C1..C12.
- Step 2.** Select “Insert” from the tool bar (Excel 2007) or the histogram-like icon in other Excel versions.
- Step 3.** Find the “Charts” group (Excel 2007), or the x-y icon in other versions of Excel.
- Step 4.** Click on the icon that shows scattered unconnected points to display a chart of the data points.
- Step 5.** On the chart just produced, right click any of the plotted points (“markers”).
- Step 6.** Click on the “Add Trendline” button.
- Step 7.** Click on the button that says “Linear.”
- Step 8.** Check the box that says “Display Equation on Chart.”
- Step 9.** Click on the “Close” button.
- Step 10.** Embellish the chart as needed by adding titles to the axes, changing the appearance of the data markers, and moving the regression equation as needed. For Example 18.2a, the finished chart should look similar to Figure 18.8.

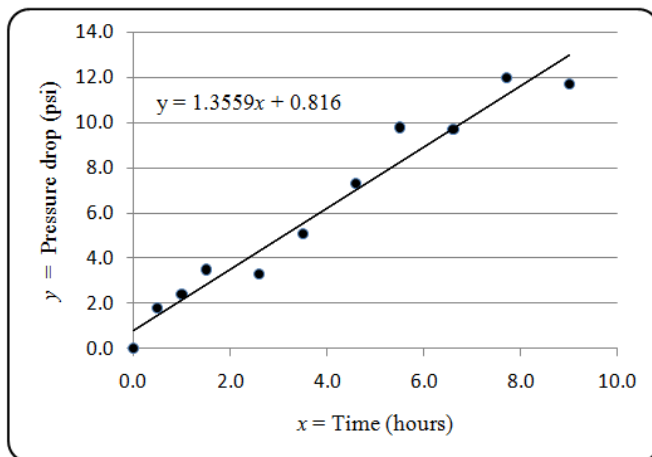


Figure 18.8. Pressure vs. time for Example 18.2a

Figure 18.8 suggests that pressure decreases over time. In the next section, we investigate whether this trend is statistically significant.

For Example 18.2b, in Step 1, select cells B1..B12 instead of A1..A12. All other steps are the same. Figure 18.9 shows the regression chart for Example 18.2b.

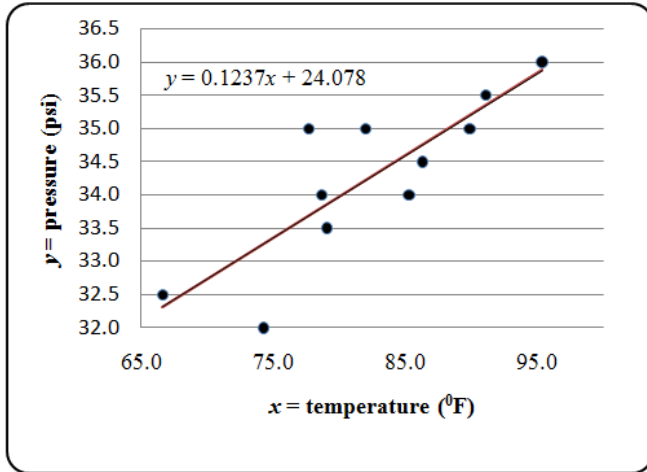


Figure 18.9. Pressure vs. temperature for Example 18.2b

Figure 18.9 suggests that pressure increases with temperature. In the next section, we investigate whether this trend is statistically significant.

18.12 Using Excel for regression analysis

To test hypotheses about the regression coefficients, use Excel's regression ANOVA routine. We illustrate the use of this routine with Examples 18.2a and 18.2b.

From Equation (18.1), the model that relates pressure to time is:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

where Y_i denotes pressure (psi) at time x_i (hours). From Equation (18.8), $\varepsilon_i \sim N(0, \sigma^2)$.

To test whether the pressure inside the containment deteriorates over time, we test the slope with the following null and (one-sided) alternative hypotheses:

$$H_0: \beta = 0 \text{ vs. } H_1: \beta < 0 \quad (18.31)$$

On very rare occasions, there is a need to test the intercept. We include this test in our example for purposes of illustration.

The null and alternative hypotheses for testing the intercept are:

$$H_0: \alpha = \alpha_0 \text{ vs. } H_1: \alpha \neq \alpha_0 \quad (18.32)$$

where α_0 is a given constant.

The regression ANOVA, unfortunately, has two drawbacks:

- The F -test that is performed by the ANOVA is a two-sided test. The ANOVA is not designed for a one-sided alternative hypothesis such as $H_1: \beta < 0$.
- The F -test can test only if a parameter is zero. Thus, the ANOVA cannot test whether the intercept specified in the null hypothesis is any value other than zero.

Excel's output, fortunately, provides some auxiliary information that allows us to overcome these drawbacks. We illustrate the use of this information in Sections 18.13 and 18.14. In this section, we run the slope analysis for a two-sided alternative hypothesis and the intercept analysis for a zero null hypothesis.

We construct the regression line and perform the analysis using Excel's regression data analysis routine, as shown in the following steps:

- Step 1.** From the tool bar, select the "Data" tab (or "Tools" in some versions of Excel).
- Step 2.** Click on "Data Analysis."
- Step 3.** Under analysis tools, scroll down and select "Regression."
- Step 4.** Next to "Input Y Range," enter the range that contains the Y values. In Table 18.7, select cells C1..C12.
- Step 5.** Next to "Input X Range," enter the range that contains the x values for Example 18.2a. In Table 18.7, we select cells A1..A12.
- Step 6.** Check the "Labels" box if variable labels are included in Steps 4 and 5, as we have here.
- Step 7.** Be sure that the box labeled "Constant is zero" is not checked.
- Step 8.** If a test of hypothesis is needed, check the "Confidence Level" box and enter the value of $100(1 - \alpha)$ next to that box. For $\alpha = 0.05$, enter the number 95.
- Step 9.** Indicate where you wish the output to print.

Step 10. Check the “Line Fit Plots” box to obtain a display of the data and the least squares line. However, the plotting procedure described in Section 18.11 is more amenable to reformatting and embellishment.

Step 11. Click “OK” to execute the routine.

We extracted the following table from Excel’s output.

Table 18.8. Regression ANOVA: Pressure vs. time (Example 18.2a)

ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	2.136	2.136	1.474	0.256	
Residual	9	13.046	1.450			
Total	10	15.182				
	<i>Coeff.</i>	<i>Std. Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	34.8285	0.5843	59.6103	5.3E-13	33.5068	36.1502
Time	-0.0076	0.0063	-1.2139	0.2557	-0.0218	0.0066

Excel’s output for Example 18.2a is explained below.

- The F -statistic in the ANOVA table has 1 degree of freedom in the numerator and $n - 2 = 9$ degrees of freedom in the denominator. The F -statistic is calculated as $f = 1.474$. As noted, the F -statistic tests if the slope of the regression line differs from zero.
- The critical value for this test is $F_{0.95}(1, 9) = 5.12$. Excel’s ANOVA table does not report this value, but it can be obtained from Table T-4 of the appendix, or calculated by the Excel function $=\text{FINV}(0.10,1,9) = 5.12$. Because $1.47 < 5.12$, we cannot conclude that $\beta \neq 0$.
- Excel reports the probability (labeled “significance F ”) of arriving at an F -value as low as the calculated $F = 1.47$ when $\beta = 0$. In this example, that probability is 0.256, clearly not smaller than $\alpha = 0.05$. Hence, we conclude that the rate of change of pressure with respect to time is not significantly different from 0. In other words, there is no statistical evidence of a trend.
- Under the heading of “Coeff.,” Excel’s output gives the intercept (34.8285) and the slope (-0.0076) of the regression line. (The slope results are in the row labeled “Time”.) Because both are estimates of the corresponding model parameters, they have associated standard errors, which Excel lists under the heading of “Std. Error.”
- The standard errors are important, as we use them (Sections 18.13 and 18.14) to test a one-sided alternative hypothesis or a null hypothesis that differs from zero.

- The ratio of a coefficient to its standard error, given under the heading of “*t* Stat,” is the *t*-statistic for testing if that coefficient equals zero.
- The *t*-statistic for the intercept tests the hypothesis $H_0: \alpha = 0$. In this example, the estimated intercept is 34.8285, the standard error of the intercept is 0.5843, the *t*-statistic is 59.6103, and the reported *P*-value is 0.000 (actually, 5.3E-13). Of course, we realize that, physically, the pressure at time 0 hours (and, for that matter, at any other time) cannot be 0.
- Excel’s ANOVA does not report the critical value for any of the regression tests. Instead, Excel provides the probability of obtaining (when the null hypothesis is correct) a statistic (be it *F* or *t*) as small a value as calculated. In the first part of the table, Excel labels this probability as “*Significance F*.” In the second part of the table, Excel labels this probability as “*P-value*.” If a reported probability is smaller than the level of significance α , the null hypothesis is rejected.
- The columns labeled “*Lower 95%*” and “*Upper 95%*” give 95% lower and upper confidence limits of 33.5068 and 36.1502 for the intercept. These columns also give 95% lower and upper confidence limits of the slope as -0.0218 and 0.0066. Note that this interval includes 0, confirming the conclusion of the ANOVA that the data do not contradict the null hypothesis of no trend in the pressure readings over time.

In parallel with the ANOVA for testing pressure versus time, we run an ANOVA for pressure versus temperature (Example 18.2b). The steps in preparation for this analysis are identical to the steps used in Example 18.2a, except that in Step 5, the “X Range” refers to the temperature readings in cells B2..B12 from Table 18.7.

Table 18.9 shows Excel’s output for Example 18.2b.

Table 18.9. Regression ANOVA: Pressure vs. temperature (Example 18.2b)

ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	10.510	10.510	20.246	0.001	
Residual	9	4.672	0.519			
Total	10	15.182				
	<i>Coeff.</i>	<i>Std. Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	24.0783	2.2760	10.5791	2.2E-06	18.9296	29.2271
Temperature	0.1237	0.0275	4.4996	0.0015	0.0615	0.1859

In contrast to Example 18.2a, the slope of the regression line (0.1237) is statistically significant. This follows from Table 18.9 where the *P*-value for the slope (temperature) is less than 0.05 or, equivalently, where the confidence

interval for the slope does not include 0. We conclude that there is a statistically significant trend in the temperature readings over time.

18.13 Hypothesis testing and confidence intervals for the slope

As noted earlier, the regression ANOVA provides only a two-sided test and is restricted to testing the null hypothesis that the parameter of interest is zero. In the next two sections, we show how to use Excel's output to relax these restrictions. Refer to Tables 18.8 and 18.9 for examples.

We begin by listing some properties of the estimator of the slope in a regression model (from Neter, et al. (1996), p. 45). The estimator B of the slope β is a random variable.

- B is a minimum variance unbiased estimator of β .
- B is normally distributed.
- The variance σ_B^2 of B (from Neter et al. (1996), p. 46) is:

$$\sigma_B^2 = \sigma^2 / S_{xx} \quad (18.33)$$

where S_{xx} is defined by Equation (18.12).

- The estimator of σ_B^2 is:

$$S_B^2 = MS_{Error} / S_{xx} \quad (18.34)$$

➤ Note that, in the ANOVA table, Excel uses the label "Residual," whereas we use the label "Error."

- The standard error of B is:

$$S_B = \sqrt{MS_{Error} / S_{xx}} \quad (18.35)$$

Using these properties, we construct the standardized statistic for testing $\beta = \beta_0$

$$T = \frac{B - \beta_0}{S_B} \quad (18.36)$$

Then T has a Student's t -distribution with $(n - 2)$ degrees of freedom.

Excel provides the t -statistic for testing $H_0: \beta = 0$ in the column labeled “ t Stat,” which is calculated by B/S_B from Equation (18.36) when $\beta = 0$.

Now suppose the null hypothesis is $H_0: \beta = \beta_0 \neq 0$. To test this, calculate the t -statistic by substituting β_0 for β in Equation (18.36). The three possible alternative hypotheses and their rejection regions are:

$$H_1: \beta \neq \beta_0. \text{ Reject } H_0 \text{ if } |T| > t_{1-\alpha/2}(n-2). \quad (18.37)$$

$$H_1: \beta > \beta_0. \text{ Reject } H_0 \text{ if } T > t_{1-\alpha}(n-2). \quad (18.38)$$

$$H_1: \beta < \beta_0. \text{ Reject } H_0 \text{ if } T < 0 \text{ and } |T| > t_{1-\alpha}(n-2). \quad (18.39)$$

We continue with Example 18.2b, where pressure is regressed against temperature. The values of x_i are given in Column A in Table 18.7, for which $\bar{x} = 82.40$, $\Sigma x^2 = 75374.00$, and using Equation (18.12), $S_{xx} = 86.64$. From Table 18.9, we have $MS_{Error} = 0.519$, and from Equation (18.34), we have $S_B = \sqrt{0.519 / 686.64} = 0.0275$. This value can also be found in the bottom row of Table 18.9 in the column “*Std. Error*” and the row “Temperature.”

Excel also provides a $100(1 - \alpha)\%$ confidence interval about β . The lower part of Table 18.9 shows the 95% lower and upper confidence limits. If β_0 does not lie between those limits and H_1 is two-sided, we reject H_0 . The statistics used for hypothesis testing can also be used to construct a confidence interval about the parameter β as follows:

$$B \pm t_{1-\alpha/2}(n-2)S_B, \text{ for a } 100(1 - \alpha)\% \text{ confidence interval} \quad (18.40)$$

If H_1 is one-sided, the t -statistic from Equation (18.36), also given at the bottom of Table 18.9, can be used as follows:

$$B - t_{1-\alpha}(n-2)S_B, \text{ for a } 100(1 - \alpha)\% \text{ lower confidence limit} \quad (18.41)$$

$$B + t_{1-\alpha}(n-2)S_B, \text{ for a } 100(1 - \alpha)\% \text{ upper confidence limit} \quad (18.42)$$

Example 18.3. Coefficient of expansion of metal rods. Suppose we wish to test a manufacturer’s claim that the expansion coefficient β of metal rods produced by his process is significantly lower than $\beta_0 = 1.23\text{E-}5 = 0.0000123$ inch/°F. From a sample of size $n = 20$ readings, we calculate $b = 8.03\text{E-}6 = 0.00000803$, $s^2 = MS_{Error} = 1.3\text{E-}6 = 0.0000013$, and $S_{xx} = 280810$.

We use a one-sided test of $H_0: \beta = \beta_0$ against the alternative $H_1: \beta < \beta_0$. If we reject H_0 , this will provide 95% assurance that the rods will not expand by more than β_0 , thus confirming the manufacturer’s claim. The test is $H_0: \beta = \beta_0$ versus $H_1: \beta < \beta_0$ at $\alpha = 0.05$. From Equation (18.35):

$$S_b = \sqrt{0.0000013 / 280810} = 0.000002152.$$

From Equation (18.36), the value of T is:

$$t = \frac{0.00000803 - 0.0000123}{0.000002151} = -1.985$$

The critical value of the test (from Table T-3 of the appendix, for $\nu = n - 2 = 18$) is $t_{0.95}(18) = 1.73$. From Equation (18.39), we reject H_0 because $-1.985 < -1.734$ and conclude that the statistical evidence supports the manufacturer's claim that the expansion coefficient is, at most, 0.0000123 inch/°F.

18.14 Hypothesis testing and confidence intervals for the intercept

As we did for the slope, we begin by listing some properties of the estimator of the intercept for a regression model.

- The estimator A of the slope α is a random variable.
- A is a minimum variance unbiased estimator of α .
- A is normally distributed.
- The variance σ_A^2 of A is (Neter et al. (1996), p. 53):

$$\sigma^2(A) = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} = \quad (18.43)$$

An alternative formula from the same source for σ_A^2 is:

$$\sigma^2(A) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right] \quad (18.44)$$

We estimate σ_A^2 by replacing σ^2 by MS_{Error} in Equation (18.44) to get:

$$S_A^2 = MS_{Error} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad (18.45)$$

- The standard error of A is:

$$S_A = \sqrt{MS_{Error} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \quad (18.46)$$

- Using these properties, we construct the standardized statistic for testing A :

$$T = \frac{A - \alpha}{\sqrt{MS_{Error} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} \quad (18.47)$$

Then T is distributed as Student's T with $\nu = (n - 2)$ degrees of freedom.

Now suppose the null hypothesis is $H_0: \alpha = \alpha_0$. To test this, calculate the t -statistic by substituting α_0 for α in Equation (18.42). We list the three possible alternative hypotheses and their rejection regions:

$$H_1: \alpha \neq \alpha_0. \text{ Reject } H_0 \text{ if } |T| > t_{1-\alpha/2}(n-2). \quad (18.48)$$

$$H_1: \alpha > \alpha_0. \text{ Reject } H_0 \text{ if } T > t_{1-\alpha}(n-2). \quad (18.49)$$

$$H_1: \alpha < \alpha_0. \text{ Reject } H_0 \text{ if } T < 0 \text{ and } |T| > t_{1-\alpha}(n-2). \quad (18.50)$$

The statistics used for hypothesis testing can also be used to construct confidence intervals about the parameter α as follows:

$$A \pm t_{1-\alpha/2}(n-2)S_A, \text{ for a } 100(1-\alpha)\% \text{ two-sided confidence interval} \quad (18.51)$$

$$A - t_{1-\alpha}(n-2)S_A, \text{ for a } 100(1-\alpha)\% \text{ lower confidence limit} \quad (18.52)$$

$$A + t_{1-\alpha}(n-2)S_A, \text{ for a } 100(1-\alpha)\% \text{ upper confidence limit} \quad (18.53)$$

As in the case of the regression slope, Excel's printout produces both hypothesis tests and confidence intervals for the intercept. In the lower part of Tables 18.8 and 18.9, we find the lower and upper confidence limits about the intercept for the level selected (the default value is 95%). For Example 18.2b, the 95% lower and upper limits for the temperature intercept are 18.93 and 22.23, respectively. This also means that a null hypothesis with a prespecified value α_0 that is not within these limits would be rejected.

18.15 Regression through the origin

zero-intercept model

Prior knowledge about the relation between x and Y may dictate that the model for the data would be a **zero-intercept model** because when $x = 0$, $Y = 0$. For example, when we buy a product or material by weight, when the weight of a product is 0, so is the cost. The zero-intercept model is written as:

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (18.54)$$

where:

Y_i = the value of the dependent variable for the i^{th} point

x_i = the value of the independent variable for the i^{th} point

β = an unknown constant (slope)

n = number of (x_i, Y_i) points

ε_i = the error associated with the i^{th} point

In analyzing the data, we usually assume that $\varepsilon_i \sim N(0, \sigma^2)$.

The least square estimate of the parameter β is (from Neter et al. (1996)):

$$B = \frac{\sum x_i Y_i}{\sum x_i^2} \quad (18.55)$$

which is an unbiased estimator of β .

regression through the origin

The least squares fit to the data, called **regression through the origin**, is:

$$\hat{Y}_i = Bx_i \quad (18.56)$$

An unbiased estimator of σ^2 is:

$$MS_{Error} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-1} = \frac{\sum (Y_i - Bx_i)^2}{n-1} \quad (18.57)$$

- Note that the denominator of Equation (18.57) has $n - 1$ degrees of freedom compared to $n - 2$ in the unrestricted model (i.e., a model that is not forced through the origin). This is because in the unrestricted model we estimate only one parameter (β) rather than two parameters (α and β).

The MS_{Error} is the mean square error in the ANOVA table. Excel gives this value under the column labeled “MS” and the row labeled “Residual.”

The structure of the ANOVA table is similar to that of Table 18.5, except for the total and error degrees of freedom.

Table 18.10. A regression through the origin ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Regression	SS_{Reg}	$v_{Reg} = 1$	$MS_{Reg} = \frac{SS_{Reg}}{v_{Reg}}$	$F = \frac{MS_{Reg}}{MS_{Error}}$
Error (residual)	SS_{Error}	$v_{Error} = n - 1$	$MS_{Error} = \frac{SS_{Error}}{v_{Error}}$	
Total	SS_{Total}	$v_{Total} = n$		

We can construct the regression through the origin line and analyze that line using Excel’s regression procedure in the data analysis nodule. This procedure is described in 11 steps in Section 18.12, except that in Step 7, we need to check the box labeled “Constant is zero.”

Example 18.4 illustrates the regression analysis for the zero-intercept model.

Example 18.4. Home size and heating cost. Table 18.11 gives the size of 10 single family homes (in ft^2) in the same neighborhood and their annual heating cost (in dollars). We know that when the home size is 0 (i.e., there is no home), the heating cost must be 0.

Table 18.11. Home size and heating cost

Home	1	2	3	4	5	6	7	8	9	10
Size (ft^2)	900	1250	1410	1449	1670	1880	2050	2135	2660	2905
Cost (\$)	1050	1535	1617	1940	1714	2020	2333	2440	2736	3012

Table 18.12 presents Excel’s ANOVA. Note the degrees of freedom for “Total” and error (labeled “Residual”) components.

Table 18.12. Excel's ANOVA for regression through the origin

ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	44554994	44554994	1575.3	1.8E-10	
Residual	9	254545	28283			
Total	10	44809539				

	<i>Coeff.</i>	<i>Std. Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
Home size	1.10	0.03	39.69	0.00	1.03	1.16

Excel leaves the line called “Intercept” in the same format as the unrestricted regular regression analysis, but leaves the intercept as 0 and all other intercept statistics as not applicable (#N/A).

As expected, there is a linear trend to the heating cost over the home size. In looking at the raw data (Table 18.11), we can see that the yearly heating cost is nearly equal to the size of the home. From the last line of the ANOVA, we see that, with 95% confidence, the annual heating cost is between 1.03 and 1.16 times the size of the house.

Regression through the origin is usually discouraged for several reasons:

- The model may not be appropriate. Thus, in Example 18.4, there may be some overhead or other fixed minimum fees that may escape our attention. If actual measurements are involved, we may have calibration issues that may be masked by forcing the regression line through the origin.
- The unrestricted least squares line is constructed to be the “best” possible fit. As such, it is better than a line that is subject to intercept restriction.

To prove the last point, we ran an ANOVA on the data in Example 18.4 without forcing the regression line through the origin, as shown in Table 18.13. In this configuration, the MS_{Error} (=18282) is nearly 35% smaller than that for the regression through the origin, for which $MS_{Error} = 28283$.

Table 18.13. Excel's ANOVA for unrestricted simple linear regression

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	3059519	3059519	167.3	1.2E-06
Residual	8	146259	18282		
Total	9	3205778			

	<i>Coeff.</i>	<i>Std. Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	336.84	138.4	2.43	0.04	17.68	656.00
Home size	0.93	0.07	12.94	0.00	0.76	1.10

18.16 Multiple linear regression

multiple linear regression In the previous sections of this chapter, we examined the relation between one dependent variable and a single independent variable. **Multiple linear regression** (MLR) examines the relation between one dependent variable and two or more independent variables.

The calculations involved in MLR are long and complex. The derivation of the regression coefficients and their analysis typically uses matrix algebra, which is beyond the scope of this book. Because Excel's regression routine offers only limited MLR support, we limit the coverage of MLR to scenarios that Excel can handle. We use Example 18.2 for illustration.

The general model for MLR is given by Equation (18.2). For two independent variables, the model is:

$$Y_i = \alpha + \beta x_i + \gamma u_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (18.58)$$

For Example 18.2, Y denotes pressure (psi) at time x (hours) and temperature u (°F). As usual, we assume that $\varepsilon_i \sim N(0, \sigma^2)$.

The data is analyzed using Excel by following the 11 steps listed in Section 18.12, except that in Step 5, the "X Range" is selected as cells A1..B12. This selection includes both time and temperature data.

Table 18.14 is a modified version of Excel's output.

Table 18.14. ANOVA for MLR with two independent variables

ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	2	11.440	5.720	12.231	0.004	
Residual	8	3.741	0.468			
Total	10	15.182				

	<i>Coeff.</i>	<i>Std. Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	24.9318	2.2435	11.1130	3.8E-06	19.7583	30.1053
Time	-0.0051	0.0036	-1.4106	0.1961	-0.0134	0.0032
Temp	0.1179	0.0264	4.4604	0.0021	0.0569	0.1788

It is useful to compare the analysis of the three-parameter model in Table 18.14 to that of the two-parameter models in Tables 18.8 and 18.9. The MLR in Table 18.14 regresses pressure simultaneously on time and temperature, while the simple linear regressions in Tables 18.8 and 18.9 regress pressure separately on time and temperature. Recall that, from Table 18.9, the effect of temperature on pressure is statistically significant, while from Table 18.8, the effect of time is not.

In the three-parameter ANOVA we have 2 degrees of freedom for regression. In the two-parameter case we have only 1 degree of freedom. Recall that the number of regression degrees of freedom is the same as the number of coefficients (excluding the intercept) in the model. The number of degrees of freedom for error (residual) in the three-parameter case is 8, compared to 9 in the two-parameter case.

- The number of error degrees of freedom in the ANOVA may be important for statistical tests, especially with a small sample. The more parameters included in the model, the smaller the number of degrees of freedom left for error.

From the column *Coeff.* in Table 18.14, the regression equation for the three-parameter case is $y = 24.932 - 0.005x + 0.118u$, where x is time and u is temperature. We compare this equation with the regression equation from Table 18.9, which is $y = 24.078 + 0.124u$. The question is whether we improve the regression fit by including time in the model, or should we just ignore time as insignificant.

Excel does not test “over and above” contributions of independent variables but we can do it ourselves by extracting the necessary entries from Excel’s output. Most commercial statistical software can do this with a single command.

We focus on the incremental contribution of time. First, we note that the regression sums of squares from the three-parameter and two-parameter cases are 11.440 and 10.510, respectively. The difference $11.440 - 10.510 = 0.930$ is attributed to the incremental contribution of time. This difference has one associated degree of freedom. The mean square for this contribution is

$0.930/1 = 0.930$. The ratio of the mean square divided by the mean square error of the three-parameter case is distributed as an F statistic with 1 and $n - 3$ degrees of freedom. In this example, F is calculated as $0.930/0.468 = 1.987$. From Table T-4 of the appendix, the critical value $f_{0.95}(1,8) = 5.32$. Since $1.987 < 5.32$, the contribution of time over and above temperature is not significant at the 0.05 level.

Excel can provide a multiple linear regression by following the 11 steps in Section 18.12 and selecting the data that correspond to the dependent variable and all the independent variables. Since the regression calculations involve matrix manipulations and matrix inversions, the number of independent variables that Excel can handle is limited.

Finally, we offer a note of caution. In running a multiple linear regression, be sure that no independent variable is a linear combination, either exactly or approximately, of one or more of the other independent variables. For example, if x_1 is temperature measured in degrees Celsius and x_2 is temperature measured in degrees Fahrenheit, Excel's routine will not work correctly. At best, the statistical package issues a warning of "colinearity," or "singularity," which means that the results are not to be trusted. At worst, the package provides meaningless or incorrect results without a warning.

18.17 Prediction

We return to Example 18.2b, where we found that temperature (x) is a significant independent variable in predicting containment pressure (Y). From Figure 18.9 and Table 18.9, the prediction equation is:

$$\hat{Y} = 24.078 + 0.124x \quad (18.59)$$

For each value of x , the prediction equation calculates a predicted value of Y . We sometimes use the notation $Y|x$ (read "Y given x ") to designate the predicted value for x . Thus, for $x = 78.7$, we obtain $Y|78.7 = 33.84$, and for $x = 85.0$, we obtain $Y|85.0 = 34.62$.

Since the x values in the dataset (in °F) from Table 18.7 are between 66.6 and 95.4, prediction should be made only for that temperature range. A prediction of pressure when the temperature is 120°F, for example, is an *extrapolation*. Extrapolations should be avoided because we have no assurance that the regression model is applicable outside the domain of the observed independent variables.

Even assuming that the regression model is correct, there are two potential sources of error in using a prediction equation to predict Y at a given value of x . First, the coefficients in the equation are estimates of the true coefficients, and

second, the equation does not account for the random error term in the model. In fact, \hat{Y} is more accurately a prediction of the expected value of Y . At best, we may expect the predicted value of Y to be near its observed value.

Since the prediction is almost certainly in error, we hedge our prediction. We do that by obtaining limits (bounds) called **prediction limits** that we are fairly sure, say, with $1 - \alpha$ probability (typically 0.95), that the true (but unknown) value of $Y|x$ is between those limits. The interval defined by these prediction limits is called a **prediction interval**.

A prediction interval may be two-sided or one-sided. A one-sided interval may be bounded by either a lower or an upper prediction limit.

Prediction limits for a two-sided $100(1 - \alpha)\%$ confidence interval for $Y|x$ are given by (see Neter, et al. (1996), pp. 61–67):

$$A + Bx \pm [t_{(1-\alpha/2)}(n-2)] S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \quad (18.60)$$

where S is the square root of the error mean square cell in the regression analysis table and S_{xx} is defined in Equation (18.12). In our example, we have $s^2 = 0.519$ (from Table 18.9), and from Table 18.7, we calculate $S_{xx} = 686.64$ and $\bar{x} = 82.40$.

Similarly, a prediction limit for a one-sided lower $100(1 - \alpha)\%$ confidence interval for $Y|x$ is:

$$A + Bx - [t_{(1-\alpha)}(n-2)] S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \quad (18.61)$$

and a prediction limit for a one-sided upper $100(1 - \alpha)\%$ confidence interval for $Y|x$ is:

$$A + Bx + [t_{(1-\alpha)}(n-2)] S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \quad (18.62)$$

*confidence
belt*

If we plot Equation (18.60), we will see two curves, one lying entirely above and the other lying entirely below the regression line. Together, these curves form a *confidence belt* that is narrowest when $x = \bar{x}$. (Although this may be hard to detect from the graph, it should be obvious from the equation.) Figure 18.10 shows a confidence belt for Example 18.2b.

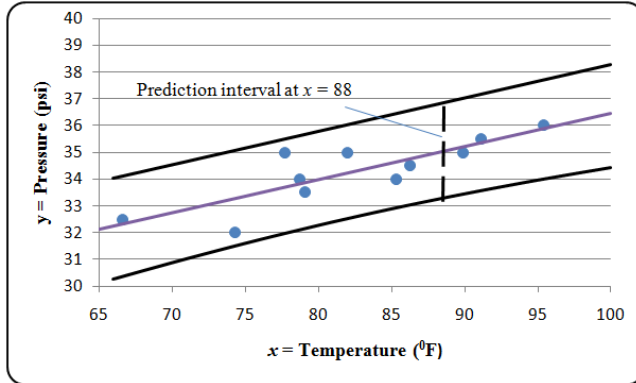


Figure 18.10. Confidence belt for Example 18.2b

From Equation (18.59), $Y|88.0 = 24.078 + 0.124(88.0) = 34.99$. From Equation (18.60), the associated two-sided prediction interval is:

$$24.078 + 0.124(88.0) \pm (2.26)\sqrt{0.519} \sqrt{1 + \frac{1}{11} + \frac{(88.0 - 82.4)^2}{686.64}}$$

which is reported, in interval form, as (33.254, 36.726).

19

Simple linear correlation

19.1 What to look for in Chapter 19

Chapter 19 provides tools to examine and quantify the relation between two normally distributed variables. In this chapter, we will learn about:

- *scatter diagram, §19.2*
- *correlation coefficient, §19.3*
- *covariance, §19.3*
- *using spreadsheet routines for correlation calculations, §19.4*
- *Fisher's Z transformation, §19.5*

We will acquire tools to test these hypotheses:

- *a correlation coefficient is zero, §19.5*
- *a correlation coefficient is a constant other than zero, §19.5*
- *two correlation coefficients are equal, §19.7*

We will learn to construct:

- *a confidence interval for a correlation coefficient, §19.6*

19.2 Basics of Simple linear correlation

scatter diagram

scatter plot

correlation coefficient

correlation, perfect

Chapter 19 examines the relationship between two normally distributed random variables, X and Y , where neither variable is necessarily a dependent or an independent variable. Given a sample of n points, $\{(x_i, y_i)\}$, we plot these points on a grid to produce a *scatter diagram* or *scatter plot* and examine it to see whether the data suggest a relation. A commonly used measure of a relationship is the *correlation coefficient*, denoted by r , a number lying between -1 and $+1$, inclusively. The variables are said to have *perfect correlation* if $r = +1$ or $r = -1$.

To gain insight into the meaning of the correlation coefficient, we consider three different types of scatter diagrams. The first scatter diagram, constructed from 50 points, is illustrated in Figure 19.1. This plot exhibits no apparent relation between the variables. In other words, knowing the value of one variable does not provide any information about the value of the second variable. This absence of a relationship is quantified by the correlation coefficient, which in this case is almost 0.

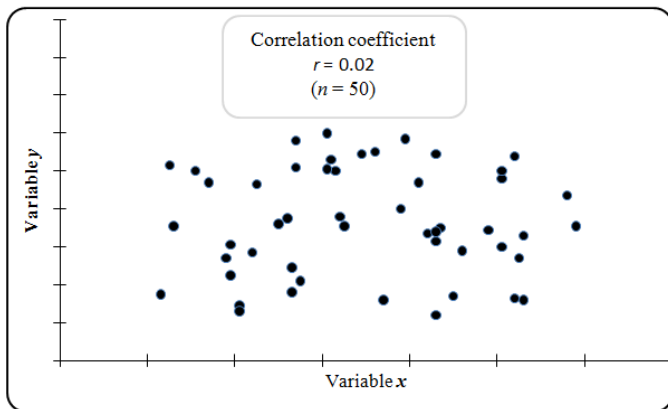


Figure 19.1. Scatter diagram with near-zero correlation

- Although the points in Figure 19.1 are a random sample from uncorrelated X 's and Y 's, the calculated correlation coefficient is not identically zero. Due to random fluctuations, very rarely will the calculated correlation be zero.

The second and third types of scatter diagrams are illustrated in Figure 19.2. Each of the two scatter diagrams exhibits a strong linear relationship between the variables. In the right-hand diagram, the points all lie close to a straight line with a positive slope (not shown). This strong linear relationship is quantified by its correlation coefficient of $r = 0.98$, almost equal to the maximum value of

+1. Had these points all fallen precisely on a straight line, they would have had perfect positive correlation. Similarly, the points in the left-hand diagram all fall close to a straight line and have an almost perfect negative correlation with $r = -0.97$.

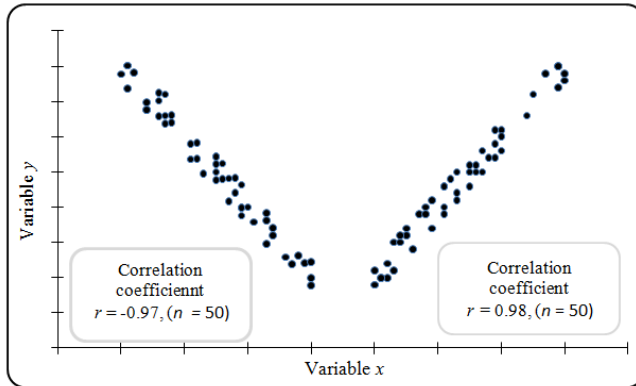


Figure 19.2. Scatter diagram for nearly perfect linear correlation

Note the adjectives “simple” and “linear” in the title of this chapter. “Simple” means that we deal with only two random variables and all our data plots have only two axes. Multi-variable correlation would involve three or more variables and investigate the simultaneous relationships among several variables or the relationship between two variables in the presence of one or more other variables.

“Linear” is used because the formula for the correlation coefficient is a measure of linearity and takes on its extreme values only when its variables are linearly related (see Section 9.3). Indeed, some writers (e.g., Gilbert (1976, p. 276)) loosely define correlation as “closeness to linearity.” This is illustrated by Figure 19.2. From Figure 9.1, we see that the absence of any pattern, i.e., non-linearity, leads to a correlation near zero. However, the converse is not true. It is possible for two variables to be functionally related and yet have zero correlation. This is illustrated in Figure 19.3, where the formula for the correlation coefficient yields $r = 0.00$. Although Y is a function of X , they are not linearly related.

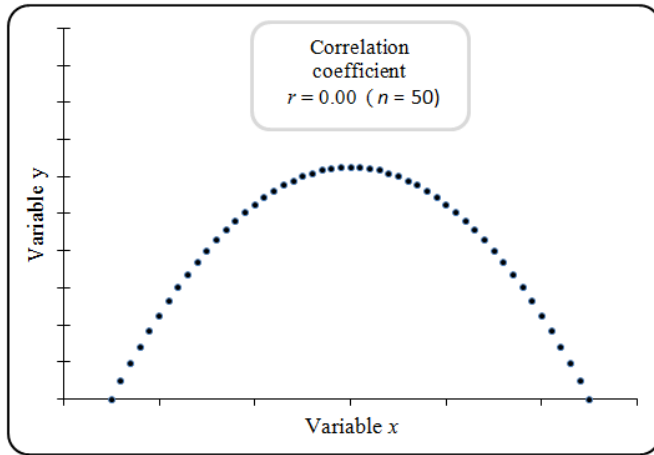


Figure 19.3. Nonlinear relationship with zero correlation

In the remainder of this chapter, we omit the adjectives *simple* and *linear* as applied to correlation.

19.3 The correlation coefficient

This section provides the necessary tools to produce the quantitative measure of the correlation between two variables. Given a sample of n points (X_i, Y_i) , we go on to construct the correlation coefficient. We start with some intermediate steps where we calculate products and cross-products of X and Y and present both definition and working formulas. The index i in these formulas runs from 1 to n .

$$S_{XX} = \sum (X_i - \bar{X})^2 \quad (\text{definition formula}) \quad (19.1)$$

$$S_{XX} = \sum X_i^2 - \frac{1}{n} \sum X_i^2 \quad (\text{working formula}) \quad (19.2)$$

$$S_{YY} = \sum (Y_i - \bar{Y})^2 \quad (\text{definition formula}) \quad (19.3)$$

$$S_{YY} = \sum Y_i^2 - \frac{1}{n} \sum Y_i^2 \quad (\text{working formula}) \quad (19.4)$$

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) \quad (\text{definition formula}) \quad (19.5)$$

$$S_{XY} = \sum X_i Y_i - \frac{1}{n} \sum X_i \sum Y_i \quad (\text{working formula}) \quad (19.6)$$

sample correlation coefficient The **sample correlation coefficient** for the two variables X and Y is given by:

$$R_{XY} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \quad (19.7)$$

Once the sample correlation coefficient R_{XY} is calculated for a dataset, it is no longer a random variable and is thus written as r_{xy} .

- The statistic R_{XY} has no units. For example, if X is measured in meters and Y in grams, both the numerator and denominator of R_{XY} have units of meters times grams and therefore cancel out.
- R_{XY} is symmetric in X and Y ; if we reverse the roles of X and Y , then $R_{XY} = R_{YX}$.
- If all x or y values are the same, we cannot calculate the sample correlation coefficient because the denominator is identically zero.
- It is easy to show that if $y_i = x_i$ for all i then $r_{xy} = 1$. Similarly, if $y_i = -x_i$ for all i , then $r_{xy} = -1$. More generally, if $y_i = \alpha + \beta x_i$ for positive β then $r_{xy} = 1$. Similarly for negative β , $r_{xy} = -1$. In other words, if Y is a linear function of X , then X and Y have perfect correlation (positive or negative).
- As stated in Section 9.2, $|R_{XY}| \leq 1$. A proof of this relation can be found in Mood et al. (1974, p. 162).
- There also is a correlation coefficient for the population (X, Y) . This coefficient is denoted by the lower-case Greek letter ρ_{XY} (read: rho-sub- XY) and satisfies all of the properties of the sample correlation coefficient listed above.
- If X and Y are independent, i.e., there is no relationship between X and Y , then $\rho_{XY} = 0$. A positive value of ρ_{XY} indicates that larger values of X tend to be paired with larger values of Y , and a negative value of ρ_{XY} indicates that larger values of X tend to be paired with smaller values of Y .
- Some writers denote the sample correlation coefficient by $\hat{\rho}$ (read: rho-hat). As introduced earlier, the “hat” symbol is used to emphasize that $\hat{\rho}$ is an estimator of the population parameter ρ_{XY} .
- Note that S_{XX} is the formula for the sum of squares of deviation about the mean for the variable X , and that $S_{XX}/(n - 1)$ is the sample variance for X , and similarly for S_{YY} .

In the remainder of this chapter, when there is no ambiguity, we omit the descriptive word *sample* from “sample correlation coefficients.”

covariance The ratio $S_{XY}/(n-1)$ is called the **covariance** between X and Y . If we change all the X 's to Y 's (or all the Y 's to X 's) in $S_{XY}/(n-1)$, we will obtain the variance of Y (or the variance of X).

The following example illustrates the calculation of the correlation coefficient.

Example 19.1. Cost of crude oil and gasoline. Table 19.1 gives the average prices of a barrel of crude oil and the retail price of a gallon of gasoline at the pump in a fictitious location over 8 consecutive years. Calculate the correlation coefficient for the two prices.

Table 19.1. Fuel cost at the source and at the pump

Year	Crude \$/barrel	Gasoline, \$/gallon	Initial calculations		
i	x	y	x^2	xy	y^2
1	19	0.82	361.00	15.5800	0.6724
2	24	0.97	576.00	23.2800	0.9409
3	31	1.15	961.00	35.6500	1.3225
4	32	1.33	1024.00	42.5600	1.7689
5	33	1.3	1089.00	42.9000	1.6900
6	28	1.15	784.00	32.2000	1.3225
7	30	1.35	900.00	40.5000	1.8225
8	23	1.29	529.00	29.6700	1.6641
Sums	$\sum x =$	$\sum y =$	$\sum x^2 =$	$\sum xy =$	$\sum y^2 =$
	220	9.36	6224.00	262.3400	11.2038

Our intermediate calculations are:

$$\text{from Equation (19.2), } S_{xx} = 6224 - (220)^2/8 = 174$$

$$\text{from Equation (19.4), } S_{xy} = 262.34 - (220)(9.36)/8 = 4.94$$

$$\text{from Equation (19.6), } S_{yy} = 11.2038 - (9.36)^2/8 = 0.2526$$

From Equation (19.7), the correlation coefficient is:

$$r_{xy} = \frac{4.94}{\sqrt{(174)(0.2526)}} = 0.7451, \text{ reportable with two significant figures as } r = 0.75.$$

A number of cautions are in order when using a correlation coefficient. For example, the interpretation of its magnitude may well depend on the application. In an exact science, such as physics or chemistry, a correlation coefficient

smaller than, say, 0.95, may suggest sloppy experimentation or irreproducible results. In other disciplines, such as psychology or medical clinical trials, a correlation coefficient as large as 0.5, say, may be evidence of an important finding.

The correlation coefficient may be the most abused and misinterpreted statistic, as it is often interpreted as demonstrating cause and effect. But correlation is not causality. To illustrate the point, take an extreme case: A study conducted at the beginning of the 20th century showed a strong correlation between the annual number of children born in Sweden and the annual number of storks reported in that country's bird sanctuaries. Should we conclude that babies are brought by storks?

Another potential pitfall is the interpretation of the magnitude of a correlation coefficient. For example, we cannot conclude that a correlation coefficient of 0.80 is "twice as good" as that of 0.40.

*intra-class
correlation*

*global
correlation*

Before calculating the correlation coefficient for a sample, be sure that the sample comes from a homogeneous population and not from two or more distinguishable sub-populations. It is in fact possible that each of several sub-samples is negatively correlated, but the combined sample is positively correlated. This phenomenon of misleading combined correlation is illustrated in Figure 19.4, where the correlation coefficient for each subsample (*intra-class correlation*) is negative, because each cluster of points leans to the upper left. In contrast, the *global correlation* coefficient which is calculated from the combined groups is positive because we have two clusters of points and those clusters lean to the right, resulting in a positive correlation.

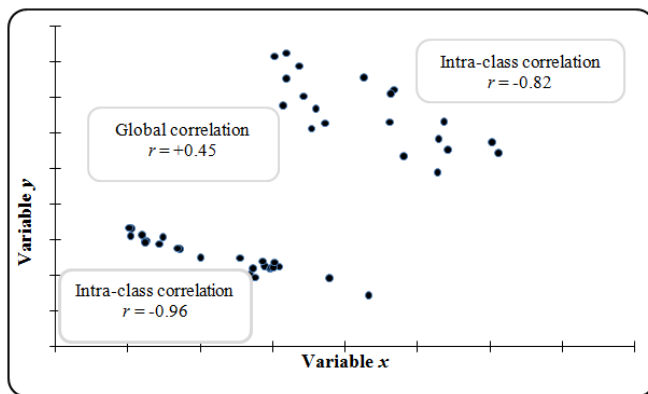


Figure 19.4. Intra-class correlation and global correlation

In the remainder of this chapter, we drop the subscript XY (or xy) of R_{XY} , r_{xy} , and ρ_{xy} , as it is obvious.

19.4 Excel's routines for calculating correlation

When there are only two variables, the correlation coefficient may be calculated by Excel's function =CORREL(range1..range2) where range1 and range2 are the locations of the first and the second variables, respectively. Note that either variable may be listed in a row or a column. For example, one variable may be listed in a column as A1,..., A11 and a second variable is listed in a row as E21,..., O21. The call =CORREL(A1..A11,E21..O21) returns the required correlation coefficient. Note that the variable name may be included as the first cell in the column (or row) of both variables. This name, however, must include non-numeric characters to assure that the name will not be included in the calculations.

When we have more than two variables, Excel offers a routine that calculates the correlation coefficients for all combinations of variables. The Excel data routine for multiple variables requires the data to be in a single array, like the one on the left side of Figure 19.5. In this example, we have four variables W , X , Y , and Z with 10 observations each in the array A1..D11. We listed variable names in the first row of the array.

	A	B	C	D	E	F	G	H	I	J
1	W	X	Y	Z			W	X	Y	Z
2	91	34	60	0.964		W	1			
3	45	77	126	0.936		X	-0.60144	1		
4	33	72	105	0.633		Y	-0.15438	-0.17627	1	
5	99	22	133	0.733		Z	0.028201	-0.2694	-0.04767	1
6	41	73	104	0.164						
7	104	39	138	0.515						
8	110	27	100	0.550						
9	98	68	68	0.666						
10	11	61	104	0.627						
11	98	73	78	0.341						

Figure 19.5. Excel's correlation input and output

Excel's correlation routine for multiple variables is conducted in the following steps:

- Step 1.** From the tool bar, select the "Data" tab (or "Tools" in some versions of Excel).
- Step 2.** Click on "Data Analysis."
- Step 3.** Under analysis tools, select "Correlation" and click "OK."

- Step 4.** Next to “Input Range,” enter the range that contains the data. In Figure 19.5 we select cells A1..D11.
- Step 5.** Next to “Group by,” select either “Columns” or “Rows.” In Figure 19.5 we select “Columns.”
- Step 6.** Check the “Labels” box if variable labels are included in Step 5. In our example, this box is checked.
- Step 7.** Indicate where you wish the output to print. In our example, we selected cell F1.
- Step 8.** Click “OK” to execute the routine.

Excel’s output gives the correlation coefficients as shown on the right side of Figure 19.5. The main diagonal of the matrix is 1, as every variable correlates perfectly with itself. The correlation matrix is symmetric about the main diagonal because the correlation between X and Y is always the same as the correlation between Y and X . However, for display purposes, Excel leaves the upper right triangle of the matrix blank.

19.5 Testing the correlation coefficient

In this section, we show how to test whether the correlation coefficient of two normal variables is equal to a pre-specified constant ρ . There are two available procedures. The first procedure uses a Student’s t -statistic that is applicable only for testing whether $\rho = 0$. The second procedure, called Fisher’s Z transformation, is more complex but it is applicable for any ρ between -1 and $+1$. Neither test requires that the mean or the standard deviation of either variable be known.

For the first procedure, the null hypothesis is:

$$H_0: \rho = 0 \quad (19.8)$$

with two- and one-sided alternatives. Student’s t -test, with $v = n - 2$ degrees of freedom is:

$$T = R \sqrt{\frac{n-2}{1-R^2}} \quad (19.9)$$

Depending on the alternative hypothesis, we reject H_0 if:

$$H_1: \rho \neq 0 \text{ and } |t| > t_{\alpha/2}(n-2) \quad (19.10)$$

$$H_1: \rho < 0 \text{ and } t < -t_{1-\alpha}(n-2) \quad (19.11)$$

$$H_1: \rho > 0 \text{ and } t > t_{1-\alpha}(n-2) \quad (19.12)$$

If H_0 is rejected, we conclude that X and Y are correlated.

Example 19.2. Testing $\rho = 0$. Suppose a sample of size $n = 10$ yields $r = 0.25$. With $H_0: \rho = 0$, $H_1: \rho \neq 0$, and $\alpha = 0.05$, we have:

$$t = 0.25 \sqrt{\frac{10-2}{1-(0.25)^2}} = 0.7303$$

Since $|t| = 0.73$ is smaller than $t_{0.975}(8) = 2.31$, we have insufficient evidence to reject the claim that $\rho = 0$. Notice, however, that the test is very insensitive to departure from $\rho = 0$. If we run a “what if” analysis, we will find that for $n = 10$, $\nu = 8$, the sample correlation coefficient must be as large as 0.64 (for which $t = 2.36$) to reject H_0 . From another point of view, a sample correlation of $r = 0.25$ would be rejected only if it were calculated from a sample of at least $n = 63$, $\nu = 61$ for which $t = 2.00$ and $t_{0.975}(61) = 2.00$, to have reached significance.

Fisher's Z transformation

To test whether the correlation coefficient differs from a non-zero constant, we use a procedure called **Fisher's Z transformation** or **Fisher's Z**. The reason for the transformation is that the sample correlation coefficient, R , is not distributed normally unless $\rho = 0$. Fisher's Z , on the other hand, does follow the normal distribution. Fisher's Z transformation is given by the expression

$$Z = \frac{1}{2} \ln \frac{1+R}{1-R} \quad (19.13)$$

The respective mean, variance, and the standard error of the random variable Z are given by:

$$E[Z] = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \quad (19.14)$$

$$\text{VAR}[Z] = \frac{1}{n-3} \quad (19.15)$$

$$\text{SE}[Z] = \sqrt{\frac{1}{n-3}} \quad (19.16)$$

where ρ is the population correlation coefficient of the normal variables X and Y .

- The mean $E(Z)$ of Z is usually denoted by the Greek lowercase letter ξ (read as “ksee”).
- An historical fact about Fisher’s Z -transformation: Mathematicians recognize it as $\text{arctanh}(R)$ or $\tanh^{-1}(R)$ (read: inverse hyperbolic tangent of R). The development of this statistic led many mathematicians and statisticians to acknowledge the genius of Fisher who first used this transformation to “normalize” the sample correlation coefficient and who derived its mean and variance.

To use Fisher’s Z effectively, we need to bring it into the standard form by dividing $[Z - E(Z)]$ by $\text{SE}(Z)$. Consistent with the notation established earlier in this book, this standardized normal should be denoted by Z . Unfortunately, the symbol Z is preempted by Fisher’s use of Z . So, for this section alone, we designate the standardized form of Z by W . We have:

$$W = \frac{\frac{1}{2} \ln \frac{1+R}{1-R} - \frac{1}{2} \ln \frac{1+\rho}{1-\rho}}{\sqrt{\frac{1}{n-3}}} \quad (19.17)$$

The null hypothesis $H_0: \rho = \rho_0$ is tested by substituting ρ_0 for ρ in Equation (19.17). Because W has a standard normal distribution if H_0 is true, we reject H_0 if:

$$H_1: \rho_0 \neq 0 \text{ and } |w| > z_{1-\alpha/2} \quad (19.18)$$

$$H_1: \rho_0 < 0 \text{ and } w < z_\alpha \quad (19.19)$$

$$H_1: \rho_0 > 0 \text{ and } w > z_{1-\alpha} \quad (19.20)$$

where w is the calculated value of W and z_q is a quantile of the standard normal from Table T-1 of the Appendix.

To illustrate the use of the W statistic, return to Example 19.2. Suppose we wish to test $H_0: \rho = 0.80$ against $H_1: \rho \neq 0.80$, and the sample correlation coefficient, based on $n = 8$ observations, is $r = 0.7451$. Complete the calculations as follows:

Step 1. Under $H_0: \rho = 0.80$, from Equation (19.14) we have:

$$E[Z] = \frac{1}{2} \ln \frac{1+0.80}{1-0.80} = \frac{1}{2} \ln(9.0) = 1.0986$$

Step 2. For a sample of size $n = 8$, from Equation (19.16) we have:

$$SE[Z] = \frac{1}{\sqrt{8-3}} = 0.4472$$

Step 3. For $r = 0.7451$, from Equation (19.13) we have:

$$z = \frac{1}{2} \ln \frac{1+0.7451}{1-0.7451} = 0.9618$$

Step 4. Putting the pieces together, the observed standardized normal variable is:

$$w = \frac{0.9618 - 1.0986}{0.4472} = -0.3059$$

Step 5. Since $-1.96 < w < 1.96$, we do not reject H_0 . We conclude that we have insufficient statistical evidence to reject the claim that $\rho = 0.80$.

In this example we used the alternative $H_1: \rho \neq 0.80$. However, in the real world, the alternative hypothesis is more likely to be one sided, such as $\rho > 0.80$ or $\rho < 0.80$. Since $|w| < 1.645$, H_0 would not be rejected with either of these alternatives.

19.6 Confidence interval for the correlation coefficient

Using procedures learned in Chapter 9, we can employ Fisher's Z -transformation to construct one- or two-sided confidence intervals for the correlation coefficient ρ . This procedure is conducted in two stages. In the first stage, we construct confidence limit(s) for the mean $E[Z]$ of Fisher's Z -transformation. In the second stage, we transform the confidence limit(s) obtained in the first stage to capture the correlation coefficient ρ with the desired confidence.

Assume that a $100(1 - \alpha)\%$ two-sided confidence interval is desired. The construction is as follows:

First stage. The endpoints of a two-sided confidence interval for $E[Z]$ are:

$$(L_1, L_2) = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \pm z_{1-\alpha/2} \sqrt{\frac{1}{n-3}} \quad (19.21)$$

where $z_{1-\alpha/2}$ is a quantile of the standard normal distribution. From Equation (19.14), a $100(1 - \alpha)\%$ two-sided confidence interval for $E[Z]$ is:

$$L_1 < E(Z) < L_2 = L_1 < \frac{1}{2} \ln \frac{1+\rho}{1-\rho} < L_2 \quad (19.22)$$

Second stage. We transform the double inequality in Equation (19.22) to a double inequality on ρ in terms of the endpoints L_1 and L_2 to obtain lower (R_1) and upper (R_2) confidence limits on ρ as:

$$R_1 = \frac{e^{2L_1} - 1}{e^{2L_1} + 1} \quad (19.23)$$

$$R_2 = \frac{e^{2L_2} - 1}{e^{2L_2} + 1} \quad (19.24)$$

For Example 19.1, we calculate a 95% two-sided confidence interval for the correlation coefficient between the costs of crude oil and premium gasoline as follows:

The relevant statistics are $r = 0.7451$, $n = 8$, and Fisher's $z = 0.9618$.

First stage. From Equation (19.21) and $\alpha = 0.05$, the endpoints for $E[Z]$ are:

$$0.9618 \pm 1.960 \sqrt{\frac{1}{5}} = 0.9618 \pm 0.8765$$

or $l_1 = 0.0853$ and $l_2 = 1.8383$.

Thus, we are 95% confident that:

$$0.0853 < \frac{1}{2} \ln \frac{1+\rho}{1-\rho} < 1.8383$$

Second stage. From Equations (19.23) and (19.24), the lower and upper confidence limits on ρ are :

$$r_1 = \frac{e^{2(0.0853)} - 1}{e^{2(0.0853)} + 1} = 0.0851 \quad \text{and} \quad r_2 = \frac{e^{2(1.8383)} - 1}{e^{2(1.8383)} + 1} = 0.9506$$

The considerable length of this interval (0.0851, 0.9506) is not very informative about the correlation coefficient, despite its seemingly “large” value of about 0.75. This result is directly attributable to the small sample size.

Example 19.3. Time and instrument drift. Instrument drift is a measure of departure from calibration due to time in service, and a significant drift, positive or negative, is worrisome. An engineer proposes to use R^2 (sometimes called the *coefficient of determination*) as a measure of correlation between instrument drift (measured in milliamps) and time (measured in months). This engineer believes that if r^2 is less than 0.30, this value is sufficient to convince us that drift does not change over time.

We have two concerns about this approach:

- $r^2 = 0.30$ may seem reasonably small, but the sample correlation coefficient is actually $r = \sqrt{0.30} = 0.55$, which may not be considered small for a correlation coefficient.
- The approach does not account for the sample size. In particular, if n is small, a confidence interval for ρ could be wide enough to include both large and small values of ρ .

To illustrate this second concern, suppose for a moment that we consider a correlation coefficient with absolute value $|\rho| < 0.30$ to be “small.” For what values of r and n can we conclude that ρ is small?

To answer this question, Table 19.2 provides two-sided 95% confidence intervals about ρ for selected values of r and n . All of the intervals have their lower limit $|r_1| < 0.30$ and their upper limit $r_2 = 0.30$, so that, with 95% confidence, we can conclude that ρ is small. Clearly, for the tabulated value of r , this conclusion would hold only for sample sizes larger than the tabulated value of n . For example, if $r = 0.25$, then n must be at least 1,316 in order to conclude that ρ is small. Even for r as small as 0.01, n must be at least 46 to be able to conclude with 95% assurance that ρ is small.

Table 19.2. Selected 95% two-sided confidence intervals

	Equation	Calculations					
<i>r</i>		0.25	0.2	0.15	0.1	0.05	0.01
<i>n</i>		1316	340	157	91	61	46
Fisher's <i>z</i>	(19.16)	0.26	0.20	0.15	0.10	0.05	0.01
<i>se</i> (<i>z</i>)	(19.16)	0.03	0.05	0.08	0.11	0.13	0.15
<i>z</i> _(1-α/2)		1.96	1.96	1.96	1.96	1.96	1.96
<i>l</i> ₁	(19.21)	0.20	0.10	-0.01	-0.11	-0.21	-0.29
<i>l</i> ₂	(19.21)	0.31	0.31	0.31	0.31	0.31	0.31
<i>r</i> ₁	(19.23)	0.20	0.10	-0.01	-0.11	-0.20	-0.28
<i>r</i> ₂	(19.24)	0.30	0.30	0.30	0.30	0.30	0.30

19.7 Testing equality of two correlation coefficients

In this section we test whether the population correlation coefficients of two independent pairs of normal variables are equal. No new tools are required for this test. Table 19.2 presents the elements of the test.

Let *Z*₁ and *Z*₂ be Fisher's *Z* for samples from two independent pairs of normal variables, (*X*₁, *Y*₁) and (*X*₂, *Y*₂), with correlations ρ_1 and ρ_2 and sample sizes *n*₁ and *n*₂, respectively. To test *H*₀: $\rho_1 = \rho_2$, we define *Z*_d = *Z*₁ - *Z*₂. Since *Z*₁ and *Z*₂ are distributed normally, so is *Z*_d. The mean *E*(*Z*_d) of *Z*_d is *E*(*Z*₁) - *E*(*Z*₂). From Equation (19.14), under *H*₀ the means of *Z*₁ and *Z*₂ are equal and *E*(*Z*_d) = 0. To find the variance of *Z*_d, recall that the variance of the difference of two independent variables is the sum of their variances (Section 6.9), so that the variance of *Z*_d is given in the bottom right-hand cell of Table 19.3.

Table 19.3. Statistics for testing equality of correlation coefficients

Sample	1	2	Difference
Correlation	<i>R</i> ₁	<i>R</i> ₂	
Fisher's <i>Z</i>	$Z_1 = \frac{1}{2} \ln \frac{1+R_1}{1-R_1}$	$Z_2 = \frac{1}{2} \ln \frac{1+R_2}{1-R_2}$	<i>Z</i> _d = <i>Z</i> ₁ - <i>Z</i> ₂
<i>E</i> (<i>Z</i>)	$E(Z_1) = \frac{1}{2} \ln \frac{1+\rho_1}{1-\rho_1}$	$E(Z_2) = \frac{1}{2} \ln \frac{1+\rho_2}{1-\rho_2}$	0, if $\rho_1 = \rho_2$
<i>V</i> (<i>Z</i>)	$V(Z_1) = \frac{1}{n_1 - 3}$	$V(Z_2) = \frac{1}{n_2 - 3}$	$V(Z_1 - Z_2) = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}$

To test H_0 , we use a statistic we call W , defined by:

$$W = \frac{\frac{1}{2} \ln \frac{1+R_1}{1-R_1} - \frac{1}{2} \ln \frac{1+R_2}{1-R_2}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \quad (19.25)$$

From Example 19.2, under H_0 , W has a standard normal distribution.

Consequently, to test the null hypothesis $H_0: \rho_1 = \rho_2$ against the alternative $H_1: \rho_1 \neq \rho_2$ with $\alpha = 0.05$, we reject H_0 if $|W| > 1.96$.

Example 19.4. Comparing correlations. The correlation between two measurement systems of water pollutants was calculated in one location from $n_1 = 80$ observations as $r_1 = 0.65$. In a second location, the correlation was measured from $n_2 = 50$ observations as $r_2 = 0.80$. We wish to test whether the correlations in the two locations are the same.

For the first location we have, from Equations (19.13) and (19.15):

$$z_1 = \frac{1}{2} \ln \frac{1+0.65}{1-0.65} = 0.7753, \text{ with variance} = \frac{1}{80-3} = 0.0130$$

Similarly, for the second location we have:

$$z_2 = \frac{1}{2} \ln \frac{1+0.80}{1-0.80} = 1.0986, \text{ with variance} = \frac{1}{50-3} = 0.0213$$

From Equation (19.25), we have:

$$w = \frac{0.7753 - 1.0986}{\sqrt{0.0130 + 0.0213}} = -1.746$$

Since $|-1.746| < z_{0.975} = 1.960$, we have no statistical evidence that the correlations at the two locations are different.

19.8 Comparison of regression and correlation analyses

Although regression and correlation analyses are based on different models and have different objectives, the two analyses have some common elements. For example, ignoring the independent-dependent characterization of the variables X and Y in a regression analysis, the correlation coefficient can, and often does, apply to regression-type data. Clearly, if it looks as though Y is linearly related

to X , we would expect that the regression slope and the correlation coefficient would both be statistically significant.

There are also several calculational similarities between correlation and regression analyses. First, both use S_{xx} and S_{yy} . Also, the coefficient of determination R^2 introduced in Example 19.3 can be calculated from the regression ANOVA table by dividing the regression sum of squares, SS_R , by the total sum of squares SS_T . In fact, the complete output of Excel's regression analysis provides R^2 (rather than R), even though we did not show it in Chapter 18.

20

Bayesian probability inference

20.1 What to look for in Chapter 20

This chapter introduces:

- *Bayesian inference and Bayes' theorem, §20.3*
- *prior and posterior distributions, §20.3*
- *Bayesian parameter estimation, §20.4*
- *conjugate priors for Poisson, exponential, and binomial data, §20.5*
- *noninformative priors for Poisson, exponential, and binomial data, §20.6*
- *Bayesian hypothesis testing, §20.8*

Some text and figures for this section were extracted and edited from materials provided to the NRC by the Idaho National Laboratory as training materials.

20.2 Motivation for Bayesian inference

The statistical estimation and inference methods introduced so far have all been exclusively based on data modeled as random samples from specified populations. In practice, this approach has two potential drawbacks. First, if the data are sparse, then the width of a confidence interval estimate of an unknown parameter may be wider than is useful for our application. Second, if we have a

model with two or more unknown parameters as inputs, there is no known way to propagate confidence interval estimates through the model to evaluate the uncertainty in the model results. An example is a probabilistic risk assessment (PRA) of a nuclear plant that calculates a core damage frequency from a logic model as a function of the reliabilities and failure probabilities of the plant structures, systems, and components.

A commonly used approach for dealing with these drawbacks is Bayesian inference. It uses information from the analyst or outside sources about the parameters that is not available from the observed data. This is the subjective probability approach discussed in Section 4.9. For parameter estimation, the additional information leads to less uncertainty than is provided by a confidence interval. For model evaluation, this approach develops distributions for the unknown parameters and propagates them through the model by simulation to yield an uncertainty distribution for the model results.

Although a theory of subjective probability has been developed based on the interpretation of probability as a degree of belief (see Laplace (1814), Jaynes (2003), and Jeffreys (1939)), it is not necessary to adopt this theory in order to use Bayesian inference. This chapter shows how to calculate posterior distributions and use them for estimation and inference.

*classical
statistical
methods*

Before applying any statistical method, it is necessary to check the assumptions on which it is based. The statistical estimation and inference methods introduced in this book, sometimes called *classical statistical methods*, are all based on a number of assumptions, as discussed in their presentation. Because it is based on the same data used for the classical methods, Bayesian inference is necessarily based on the same assumptions. While Bayesian inference claims to reduce the uncertainty associated with the classical statistical methods described in this book, this approach has two additional sources of uncertainty. The first stems from uncertainty about the chance of occurrence of an event in the real world based on someone's subjective judgment. The second stems from the diversity of results when more than one subjective judgment is elicited (e.g., when a panel is used in an expert elicitation process). Although these uncertainties should be evaluated before applying Bayesian inference, their consideration is beyond the scope of this book.

20.3 Bayesian inference

*Bayesian
inference*

*Bayes'
Theorem*

Bayesian inference is a process, using *Bayes' Theorem*, of combining data with additional information from an analyst's experience and judgment and/or outside expert opinion to estimate parameters and evaluate the plausibility of hypotheses.

Bayesian inference starts with three elements: a hypothesis H , data D , and relevant information X . H is a statement about a state of nature or about a parameter, D comes from the random sample from a population determined by H , and X is prior information about H from outside experts or the analyst's experience or judgment before D is observed. Using Bayes' Theorem, these three elements are combined to calculate the probability of H based on both D and X . This probability is the basis for Bayesian inference.

As discussed in Section 4.8, Bayes' Theorem provides a formula for the conditional probability of H given D in terms of the conditional probability of D given H and the probabilities of H and D . From Equation (4.24), this can be written as:

$$Pr\{H | D\} = Pr\{H\} \frac{Pr\{D | H\}}{Pr\{D\}} \tag{20.1}$$

Bayesian inference departs from the classical approach by explicitly using the additional information X to determine the probabilities in Equation (20.1). This results in the modified version of Equation (20.1) shown in Figure 20.1.

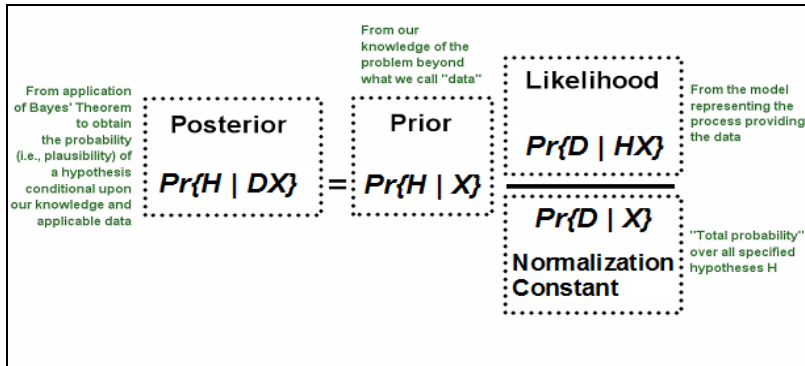


Figure 20.1. Bayes' Theorem for Bayesian inference

In Figure 20.1 the four terms are all labeled in accordance with the interpretation of Bayes' Theorem for Bayesian inference. Note that all of the probabilities are conditioned on X .

- prior** The **prior** is the probability of H before D is observed, and is based only on X . Priors can be broadly classified as either informative or noninformative. An **informative prior**, as the name suggests, contains substantive information about H .
- informative prior**
- non-informative prior** A **noninformative prior**, on the other hand, is intended to be in some sense objective and let the data speak for themselves; thus, it contains little or no substantive information about H . Other

prior distribution	terms for noninformative priors are diffuse priors, formal priors, vague priors, flat priors, and reference priors. The set of priors over all possible values of H is called the prior distribution .
likelihood	The likelihood is the probability of observing the data if H is true and is calculated from an assumed model for D . In PRA applications, the model is most often binomial, Poisson, or exponential.
posterior distribution	The posterior is the updated probability of H after D has been observed and is the basis for Bayesian inference. The set of posteriors over all possible H is called the posterior distribution .
normalization constant	The normalization constant is the unconditional probability of D . It is calculated by summing the numerator (product of the prior and likelihood) over all possible H and guarantees that the sum of all possible posteriors is 1.

Bayes' Theorem in Equation (20.1) and Figure 20.1 is written in terms of probabilities when D comes from a discrete distribution. However, it also applies when D comes from a continuous distribution and the probabilities are replaced by densities. For the continuous case, Bayes' Theorem is written as:

$$\pi_1(\theta | y) = \frac{f(y | \theta)\pi(\theta)}{\int f(y | \theta)\pi(\theta)d\theta} \quad (20.2)$$

In Equation (20.2), H has been replaced by a parameter, denoted by θ , and D is replaced by y . Corresponding to the discrete case, $\pi(\theta)$ is the prior distribution of θ , the likelihood $f(y | \theta)$ is the density of y as a function of θ , and $\pi_1(\theta | y)$ is the posterior distribution of θ . The denominator in Equation (20.2) is equal to $f(y)$, the marginal or unconditional distribution of y . Note that it is a weighted average of the likelihood, with the prior distribution for θ as the weighting function.

20.4 Bayesian parameter estimation

Let Y be a discrete random variable with an unknown parameter θ that can take on only discrete values $\theta = \theta_1, \theta_2, \dots$. Let $H = \{\theta = \theta_i\}$ be a statement about the specific value of θ , and let $Pr\{\theta = \theta_i\}$ be a prior distribution for θ . Given an observation y from Y , from Equation (20.1), the posterior is:

$$Pr\{\theta = \theta_i | Y = y\} = \frac{Pr\{Y = y | \theta = \theta_i\}Pr\{\theta = \theta_i\}}{\sum_j Pr\{Y = y | \theta = \theta_j\}Pr\{\theta = \theta_j\}} \quad (20.3)$$

If Y is continuous with density $f(y)$, then Equation (20.3) becomes:

$$g_{post}(\theta_i) = \frac{f(y|\theta_i)g_{prior}(\theta_i)}{\sum_j f(y|\theta_j)g_{prior}(\theta_j)} \tag{20.4}$$

credible interval The posterior for θ in Equation (20.3) or (20.4) is a discrete distribution. An interval of values for θ from the posterior distribution is called a **credible interval** for θ .

20.5 Conjugate distributions

conjugate prior, conjugate Given a likelihood, the formula for calculating a posterior from a prior is given by Equation (20.2), (20.3), or (20.4). While this formula can be evaluated by calculation (e.g., by a spreadsheet, numerical integration, or simulation), it is often convenient to have a closed form formula for the posterior. If the posterior can be written in closed form, then the prior is called a **conjugate prior** and is said to be **conjugate**. Whether a prior is conjugate depends not only on the form of the prior but also on the form of the likelihood.

It is possible for a posterior to have the same functional form as that of a conjugate prior. In this case, the interpretation of Bayesian inference is quite straightforward: the data updates the prior to the posterior by simply updating the parameters of the prior.

Poisson data. For modeling initiating event frequency, failure to run, or standby unavailability, the likelihood is typically a Poisson distribution (see Chapters 8 and 23). If Y is Poisson with parameter λt , the probability of y occurrences in a specified time t is:

$$f(y|\lambda) = \frac{(\lambda t)^y e^{-\lambda t}}{y!}, \quad y = 0, 1, \dots \tag{20.5}$$

gamma distribution For a Poisson likelihood, the conjugate prior turns out to be a **gamma distribution**, denoted by $Gamma(\alpha, \beta)$, with parameters $\alpha > 0$ and $\beta > 0$. For a gamma prior, the prior probability of y is given by:

$$f(y) = \frac{\beta^\alpha y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} \tag{20.6}$$

for $y > 0$, where $\Gamma(\alpha)$ is the gamma function. The gamma distribution has mean α/β and variance α/β^2 .

➤ In the gamma distribution for y , α/β has the same units as y .

The gamma distribution is discussed in Section 7.14. Figure 7.18 shows several examples of gamma distributions for $\beta = 1$.

Let Y be the number of occurrences in a specified time t . Suppose Y is Poisson with parameter λt , where λ is units of occurrences per unit time. Assume that the prior for λ is a gamma distribution, $Gamma(\alpha_{prior}, \beta_{prior})$. Suppose that y occurrences are observed in the time period t . Then the posterior also has a gamma distribution, $Gamma(\alpha_{post}, \beta_{post})$, with:

$$\alpha_{post} = \alpha_{prior} + y \quad (20.7)$$

$$\beta_{post} = \beta_{prior} + t$$

The prior can be interpreted as having observed a number of occurrences α_{prior} over a prior exposure time β_{prior} before the y occurrences were observed. It follows from Equation (20.7) that the posterior can be interpreted as observing a total number of occurrences α_{post} over a total exposure time β_{post} .

The mean of the posterior distribution is:

$$E[\lambda] = \frac{\alpha_{prior} + y}{\beta_{prior} + t} \quad (20.8)$$

and the variance is:

$$V[\lambda] = \frac{\alpha_{prior} + y}{(\beta_{prior} + t)^2} \quad (20.9)$$

Quantiles of the posterior distribution can be obtained using the =GAMMAINV(quantile, α , $1/\beta$) function in Excel. Note that Excel uses $1/\beta$ as the second parameter in =GAMMAINV().

Jeffreys prior For purposes of Bayesian updating, the noninformative **Jeffreys prior** for Poisson data is a gamma distribution with the first parameter equal to 0.5 and the second parameter equal to 0. This can be interpreted as half an occurrence in zero time. Although this prior is an improper distribution, from Equation (20.7), it leads to a proper posterior distribution, $Gamma(y + 0.5, t)$. Another noninformative prior, useful with sparse data, is an improper gamma distribution with both parameters equal to zero. This yields a proper posterior distribution $Gamma(y, t)$, provided $y > 0$.

Example 20.1. Loss of offsite power. Assume the prior distribution for loss of offsite power (LOOP) frequency is $Gamma(1.737, 52.47)$. The prior mean is $1.737/52.47 = 0.033$ occurrences per year. A plant experiences one LOOP in 9 years. Find the posterior mean and a 90% credible interval for the LOOP frequency at that plant.

The posterior distribution is $Gamma(1.737 + 1, 52.47 + 9)$. The posterior mean is therefore $2.737/61.47 \text{ yr} = 0.045/\text{yr}$. This is larger than the prior mean of $0.033/\text{yr}$ because the sample mean of the observed data is $1/9 = 0.11/\text{yr}$, and the posterior mean is a weighted average of the prior and sample means.

A symmetric 90% credible interval is bounded by the 0.05 and 0.95 quantiles of the posterior. The 0.05 quantile is found using $=GAMMAINV(0.05, 2.737, 1/61.47) = 0.011/\text{yr}$. Similarly, the 0.95 quantile is found by changing the first argument of $=GAMMAINV()$ to 0.95, yielding $0.096/\text{yr}$. Thus, a 90% credible interval for the LOSP frequency is $(0.011, 0.096)/\text{yr}$.

Exponential data. In PRA applications, exponential data will usually arise when the random variable being observed is time (e.g., time to failure, time to repair, time to recover from LOSP, or time to suppress a fire). The observed data are t_1, t_2, \dots , where t_1 is the time to the first event, t_2 is the time between the first and the second event, and so on. We assume that the intervals between events are independent and have an exponential distribution with parameter λ (see Section 7.13 Here, λ is a frequency, with units of occurrences per unit time.

If n events are observed, the likelihood is:

$$f(t_1, t_2, \dots, t_n | \lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^n t_i). \quad (20.10)$$

The conjugate prior is again a gamma distribution, just as for Poisson data. The posterior distribution will also be gamma, with parameters $(\alpha_{prior} + n)$ and $(\beta_{prior} + \sum t_i)$. From Equation (20.8), the mean of the posterior is:

$$E[\lambda] = \frac{\alpha_{prior} + n}{\beta_{prior} + \sum_{i=1}^n t_i} \quad (20.11)$$

For purposes of Bayesian updating, the Jeffreys prior for exponential data is an improper gamma distribution with both parameters equal to 0, yielding a proper posterior of $Gamma(n, \sum t_i)$.

Example 20.2. Cable fires. We investigate the suppression rate for cable fires. Assume that the prior distribution for the suppression rate is a gamma distribution with the first parameter equal to 0.5 and a mean value of 0.2 minutes. Suppose that the following times (in minutes) to suppress cable fires have been observed: 2, 2.5, 5, 10. We wish to find the posterior mean and a 90% credible interval for the suppression rate.

We first have to find the second parameter of the gamma prior. We know that the mean is the ratio of the first parameter to the second, so we have $0.5/\beta_{prior} = 0.2$. Therefore, $\beta_{prior} = 0.5/0.2 = 2.5$ minutes. Thus, the prior is equivalent to data of 0.5 cable fires in 2.5 minutes.

The observed data are four fires in a total of 19.5 minutes. Therefore, the posterior distribution is a gamma distribution with parameters $0.5 + 4 = 4.5$ and $2.5 \text{ min} + 19.5 \text{ min} = 22 \text{ min}$. Therefore, the posterior mean suppression rate is $4.5/22 \text{ min} = 0.20/\text{min}$. The 5th percentile is found using =GAMMAINV(0.05, 4.5, 1/22) = 0.08/min. Similarly, the 95th percentile is found by changing the first argument of =GAMMAINV() to 0.95, yielding 0.38/min. Thus, a 90% credible interval for the posterior suppression rate is (0.08, 0.38)/min.

Note that the prior and posterior means are the same to two decimal places. This happens because the prior mean and the sample mean are nearly equal. However, a 90% credible interval for the prior is significantly broader than the corresponding posterior interval, so the uncertainty in the suppression rate has been reduced by the observed data.

Binomial data: For modeling failure to change state on demand (e.g., failure to start), we typically assume the number of failures in a specified number of demands is a binomial random variable Y with parameters n and p (see Section 8.7). The likelihood is the binomial distribution:

$$f(y | p) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1, \dots, n \quad (20.12)$$

For the binomial likelihood, the conjugate prior is a beta distribution with parameters α and β :

$$f(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \quad (20.13)$$

for $\alpha > 0, \beta > 0, 0 \leq p \leq 1$, and where $Beta(\alpha, \beta)$ is the beta function defined by:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx, \quad \alpha, \beta > 0 \quad (20.14)$$

Another form of the beta function, useful for calculation, is:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (20.15)$$

From Equation 7.65, the mean of a beta distribution is the first parameter divided by the sum of the two parameters. Hence,

$$E[p] = \frac{\alpha}{\alpha + \beta} \quad (20.16)$$

From Equation 7.66, the variance of a beta distribution is:

$$V(p) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (20.17)$$

As a prior, the beta distribution is equivalent to data, with α equal to the prior number of failures and $\alpha + \beta$ equal to the prior number of demands.

Figures 7.19 and 7.20 give examples of the beta distribution for selected α and β .

Assume a beta prior with parameters α_{prior} and β_{prior} for p and data of y failures in n demands and a binomial likelihood. The posterior distribution of p is also a beta distribution with parameters:

$$\begin{aligned} \alpha_{post} &= \alpha_{prior} + y \\ \beta_{post} &= \beta_{prior} + n - y \end{aligned} \quad (20.18)$$

Consistent with the interpretation of the prior, the posterior is equivalent to data with α_{post} failures in $\alpha_{post} + \beta_{post}$ demands.

Quantiles of the posterior distribution can be obtained using the =BETAINV function in Excel, =BETAINV(quantile, α , β).

For binomial data, the Jeffreys noninformative prior is a beta distribution with both parameters equal to 0.5. Therefore, the posterior is $Beta(y + 0.5, n - y + 0.5)$. The posterior mean is $(y + 0.5)/(n + 1)$, a formula sometimes used in PRA applications for parameter estimation. Another noninformative prior, which can be useful with sparse data, is a $Beta(0, 0)$ distribution. This is an improper prior, but if $y > 0$, the posterior will be a proper beta distribution with parameters y and n , and the posterior mean is y/n .

Example 20.3. Emergency diesel generators. Assume an emergency diesel generator (EDG) has a prior distribution for demand failure that is $Beta(0.957, 190)$. Over the past evaluation period, there have been 26 valid demands on the EDG, with no observed failures. We wish to find the posterior distribution for the demand failure probability, the posterior mean, and a 90% posterior credible interval.

The posterior distribution is a beta distribution with parameters $0.957 + 0 = 0.957$ and $190 + 26 - 0 = 216$. Hence, the posterior mean is $0.957/(0.957 + 216) = 0.0044$.

The 5th percentile is found by $=BETAINV(0.05, 0.957, 216) = 2 \times 10^{-4}$. Similarly, the 95th percentile is found by changing the first argument of $=BETAINV$ to 0.95, yielding 0.013.

20.6 Noninformative prior distributions

As the name indicates, a prior distribution is supposed to be chosen before the data are observed. While the analyst often has some prior information about a probability or parameter of interest, it may not be complete enough to specify a prior distribution. In such cases, it may be tempting to wait until the data are observed to choose a prior, but this would contradict the basic premise that a prior should be chosen before the data are observed. If the prior is chosen or is modified after the data are observed, then the data are being given more weight than they deserve.

If a prior distribution based only on prior information cannot be specified, then a noninformative prior may be used. As defined in Section 20.3, a noninformative prior contains little or no substantive information about the probability or parameter specified by the hypothesis. How can such a prior be chosen? Reverend Bayes suggested a uniform prior and Laplace used this, but such a prior presents philosophical and mathematical problems.

As an alternative, Jeffreys suggested a prior that was insensitive to variations in scale and location. Consequently, such a noninformative prior is typically not uniform. Instead, it depends on the process that generated the data and also has the property that the Bayesian credible intervals are approximately equal to

classical confidence intervals. The Jeffreys noninformative priors for the data models introduced in Section 20.5 are as follows:

- Poisson data: Noninformative prior for λ is $Gamma(1/2, 0)$
- Exponential data: Noninformative prior for λ is $Gamma(0, 0)$
- Binomial data: Noninformative prior for p is $Beta(1/2, 1/2)$

Figure 20.2 compares 90% confidence intervals with 90% credible intervals for p for three sets of binomial data: 0 failures in 5 demands, 1 failure in 10 demands, and 3 failures in 30 demands. The credible intervals are shorter than the confidence intervals, but they become more similar as the amount of data increases. This occurs because the influence of a prior on the posterior decreases as the amount of data increases.

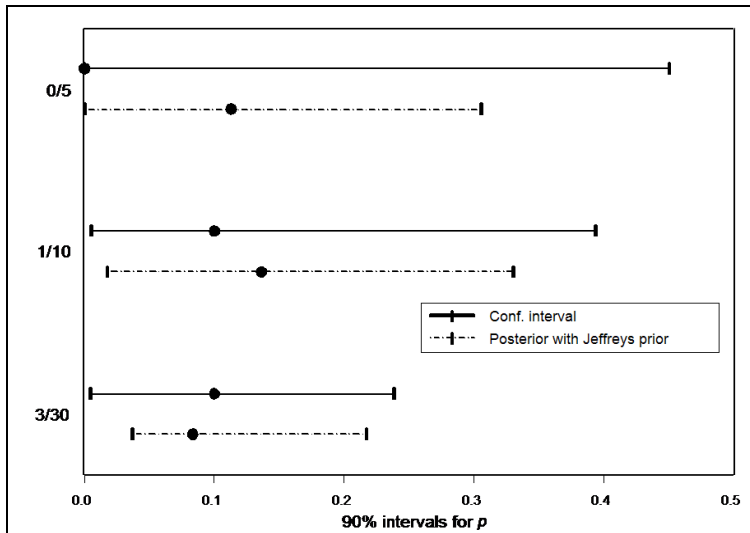


Figure 20.2. Confidence and credible intervals for binomial data

20.7 Nonconjugate prior distributions

If posterior distribution is $Gamma(\alpha, \beta)$ or $Beta(\alpha, \beta)$ with α small (much smaller than 0.5), then the lower tail of the posterior may be unrealistically small. For this or other reasons, we may prefer to use a nonconjugate prior. When the prior is not conjugate, the posterior distribution must be found by numerical integration or by simulation.

Sometimes we may have a prior, such as a nonconjugate prior, that we prefer not to use. Instead, we may select a “similar” prior of a different, but more convenient, functional form (such as a conjugate prior). To replace one prior with another, the easiest way in most cases is to adjust them to have the same

mean and variance. This matching approach can usually be done with algebra alone. Other approaches, such as matching means and some specific percentile (e.g., the 95th), are typically harder to use.

- Note that “similar” priors do not necessarily have similar percentiles, or produce posterior distributions with similar percentiles.

A random variable X has a lognormal distribution if $\ln(X)$ is normal. Section 7.9 describes the properties and characteristics of the lognormal. The lognormal distribution is often used as a prior distribution in reliability estimates, even though it is not conjugate. As with the general case of nonconjugate priors, either numerical integration or simulation (see Chapter 27) may be applied to construct the posterior or a conjugate distribution may be fitted to the lognormal prior (Hamada, et al. (2008)).

20.8 Bayesian hypothesis testing

We illustrate Bayesian hypothesis testing with an example.

Example 20.4. EDG performance. We wish to have reasonable assurance that a specific EDG will start and load with a probability of 0.975 or greater. Accordingly, we test a hypothesis about the EDG failure probability, denoted by p . The null and alternative hypotheses are:

$$H_0: p \leq 0.025, \text{ and } H_1: p > 0.025 \quad (20.19)$$

In the Bayesian framework, we start with a prior distribution for p , probably based on generic data from similar EDGs. Accordingly, we assume that the prior distribution is $Beta(0.92, 17.55)$. From Excel, the mean is 0.05 and a 90% credible interval is (0.002, 0.15).

The probability that H_0 is true (i.e., $p \leq 0.025$) is given by the area under the prior distribution to the left of 0.025. This can be found using the =BETADIST function in Excel. Applying this Excel function results in a probability of 0.40.

We next collect data related to p , making our usual assumption that the number of failures, y , in n demands follows a binomial distribution with parameters n (specified) and p (unknown). Assume that we see two failures in 15 demands on the EDG to start and load.

Because we started with a beta prior and assumed a binomial model for the likelihood, the posterior distribution will also be beta, with parameters $\alpha_{post} = \alpha_{prior} + y = 0.92 + 2 = 2.92$ and $\beta_{post} = \beta_{prior} + n - y = 17.55 + 15 - 2 = 30.55$. The posterior mean = $\alpha_{post}/(\alpha_{post} + \beta_{post})$, and the 5th percentile is found using =BETAINV(0.05, α_{post} , β_{post}), yielding 0.024/yr. Similarly, the 95th percentile is found by changing the first argument of =BETAINV to 0.95, yielding 0.178/yr. A 90% credible interval is thus (0.024, 0.178). The mean is

0.087. Both the posterior mean and 90% credible interval bounds are larger than the corresponding prior values.

From Excel, the posterior probability that H_0 is true is 0.05. Therefore, the Bayes update has decreased the probability (from 40% to 5%) that the EDG meets its performance criterion of less than 2.5 failures in 100 demands.

20.9 Bayes factors

Bayes factor Bayesians often use what is called a **Bayes factor** to help make decisions. The Bayes factor is defined in terms of odds. The **odds of an event** (or a hypothesis) is defined as the ratio of the event probability to the complement of the event probability; that is, odds = event probability/(1 – event probability). The Bayes factor against H_0 , denoted by B_f , is defined as the ratio of the posterior odds against H_0 to the prior odds against H_0 . In other words, B_f measures how the odds against H_0 have been changed by the observed data. In general, B_f depends on both the data and the prior distribution. Note that the Bayes factor is defined only for proper prior distributions; if the prior is improper (e.g., a Jeffreys prior for Poisson data), a Bayes factor cannot be calculated.

For Example 20.4, the prior odds against H_0 are:

$$Pr\{H_0 \text{ false} \mid \text{prior}\} / Pr\{H_0 \text{ true} \mid \text{prior}\} = 0.602/0.398 = 1.51.$$

The posterior odds against H_0 are:

$$Pr\{H_0 \text{ false} \mid \text{posterior}\} / Pr\{H_0 \text{ true} \mid \text{posterior}\} = 0.947/0.053 = 17.8.$$

The Bayes factor against H_0 is therefore:

$$B_f = 17.8/1.51 = 11.8.$$

Once we have calculated a Bayes factor, how do we use it to make a decision? Harold Jeffreys (Jeffreys, 1961) suggested the following table to relate the Bayes factor to the strength of the evidence against H_0 .

Table 20.1. Bayes factors and the strength of evidence

Values of B_f	Evidence against H_0
$1 < B_f < 3$	Weak
$3 < B_f < 12$	Positive
$12 < B_f < 150$	Strong
$B_f > 150$	Decisive

For Example 20.4, B_f is 11.8, so we have positive evidence against H_0 . Hence, the EDG might warrant further investigation and closer monitoring during the next evaluation period.

20.10 Consistency of the prior with the observed data

Example 20.4 focused on testing whether p exceeds a specific value. We may also be interested in whether the distribution of p appears to be consistent with the observed equipment performance.

Example 20.5. Tossing a coin. Suppose someone we know offers to toss a coin. Our knowledge of the person leads us to believe there is a significant chance that the coin is two-headed. (For simplicity, we ignore the possibility of a two-tailed coin.) Accordingly, we have two hypotheses:

$$H_1: \text{the coin is fair, and } H_2: \text{the coin is two-headed} \quad (20.20)$$

The first step in the Bayesian approach is to assign a prior to the probabilities of the hypotheses. Assume the following prior probabilities for the two hypotheses:

$$Pr\{H_1\} = 0.75, \text{ and } Pr\{H_2\} = 0.25 \quad (20.21)$$

The prior belief about the probabilities of the two hypotheses can be expressed in terms of odds. From Equation (20.21), the odds that the coin is fair are 3:1.

Now suppose the coin is tossed once, and it comes up heads. Setting $D = \{\text{heads}\}$, the likelihood is:

$$p_i = Pr\{D | H_i\} \quad (20.22)$$

where $p_i = Pr\{\text{the coin comes up heads on a single toss conditional on } H_i\}$, $i = 1, 2$.

From Equation (20.20):

$$p_1 = 0.5, p_2 = 1 \quad (20.23)$$

The normalization constant $Pr\{D\}$ in Figure 20.1 is found by summing the prior probabilities multiplied by the likelihoods over all possible hypotheses. We have:

$$Pr\{D\} = Pr\{H_1\}p_1 + Pr\{H_2\}p_2 = 0.75(0.5) + 0.25(1.0) = 0.625 \quad (20.24)$$

From Equation (20.1), the posterior probabilities are then:

$$Pr\{H_1 | D\} = 0.6, Pr\{H_2 | D\} = 0.4 \tag{20.25}$$

Table 20.2 presents the results after one toss.

Table 20.2. Bayesian inference after one toss in Example 20.5

Hypothesis	Prior Probability	Likelihood	(Prior) × (Likelihood)	Posterior Probability
H ₁ : fair coin (i.e., Pr{head} = 0.5)	0.75	0.5	0.375	0.60
H ₂ : two-headed coin (i.e., Pr{head} = 1.0)	0.25	1.0	0.250	0.40
Sum	1.00		0.625	1.00

As a result of tossing a single head, the odds that the coin is fair have decreased by a factor of 2 from 3:1 to 3:2. Now suppose that we continue tossing the coin and keep getting heads. Figure 20.3 shows the results for 10 tosses. After five tosses, the probability that the coin is fair is less than 0.1 (i.e., the odds are about 10:1 against having a fair coin). After 10 tosses, the odds against a fair coin are extremely large.

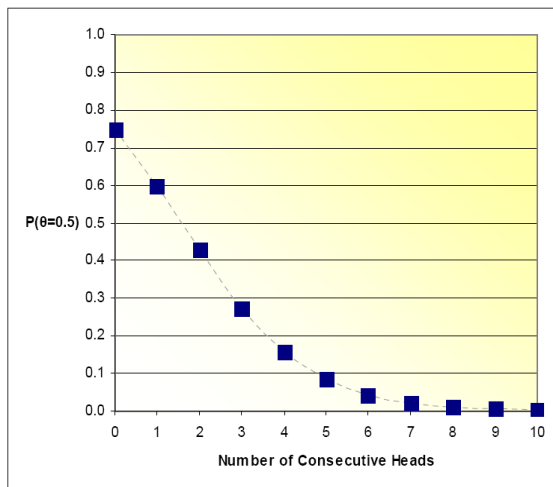


Figure 20.3. Posterior probability of a fair coin as a function of the number of consecutive heads

21

Hypergeometric experiments

21.1 What to look for in Chapter 21

Chapter 21 deals with the hypergeometric distribution, a distribution that applies to sampling without replacement. The hypergeometric distribution was already introduced in Section 8.6; this chapter illustrates its applications.

In our coverage of the hypergeometric distribution, we are reminded of several terms and concepts already covered in Chapter 8:

- *attributes, §21.2*
- *hypergeometric distribution function, §21.2*
- *mean and variance of the hypergeometric distribution, §21.2*
- *sampling without replacement, §21.2*

In this chapter, we learn to:

- *calculate hypergeometric probabilities, §21.2*
- *estimate the proportion and number of items with attributes, §21.3*
- *test hypotheses about the number of items with attributes, §21.4*
- *choose a sample size for a hypothesis test, §21.5*
- *approximate the hypergeometric by a normal distribution, §21.6*

21.2 Basics of the hypergeometric distribution

hypergeometric distribution The *hypergeometric distribution* is a discrete distribution that calculates the probability of selecting a random sample of items with a specific (binary) **attribute** from a finite population. The distinguishing feature of the distribution is that it applies to sampling without replacement experiments.

success Most statistical books refer to an item with an attribute, be it in the population or in the sample, as a *success*. There is no implication that a success is good or that it is bad.

From Section 8.6, the probability function of the hypergeometric distribution is:

$$\begin{aligned}
 f(y) &= f(y; N, M, n) = \\
 &= \Pr\{y/N, M, n\} = \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}, \quad (21.1) \\
 & \quad y = \max(0, n + M - N), \dots, \min(M, n)
 \end{aligned}$$

where:

N = population size (finite)

M = number of successes in the population

n = sample size

y = number of successes in the sample

$\Pr\{y\} = \Pr\{y/N, M, n\}$ = probability of obtaining exactly y successes, given N, M, n , and:

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} \text{ is the binomial coefficient, defined in Section 8.6.}$$

➤ Section 8.6 explains the minimum and maximum limits on y .

Example 21.1. Hypergeometric probability calculations. The calculations involved in Equation (21.1) are illustrated for a sample of size $n = 5$ for a population of $N = 50$ items that includes $M = 10$ defective items. Calculated manually, the probability that the sample contains $y = 1$ defective items is:

$$Pr\{1 | 50, 10, 5\} = \frac{\binom{10}{1} \binom{50-10}{5-1}}{\binom{50}{5}} = \frac{10!}{1! 9!} \frac{40!}{4! 36!} \frac{5! 45!}{50!} = 0.4313$$

The hypergeometric probability calculations in this chapter are made by calling Excel's function =HYPGEOMDIST(y, n, M, N). In this example =HYPGEOMDIST(1, 5, 10, 50) returns the value of 0.4313.

From Section 8.6, the mean, variance, and standard deviation of the hypergeometric distribution Y are:

$$E[Y] = \mu = \frac{nM}{N} \quad (21.2)$$

$$V[Y] = \sigma^2 = \frac{nM}{N} \frac{N-M}{N} \frac{N-n}{N-1} \quad (21.3)$$

$$SD[Y] = \sigma = \sqrt{\frac{nM}{N} \frac{N-M}{N} \frac{N-n}{N-1}} \quad (21.4)$$

Using Equations (21.1) - (21.4) for Example 21.1, we calculate:

$$\mu = \frac{(5)(10)}{50} = 1.00$$

$$\sigma^2 = \left(\frac{(5)(10)}{50} \right) \left(\frac{50-10}{50} \right) \left(\frac{50-5}{50-1} \right) = 0.735$$

$$\sigma = \sqrt{0.735} = 0.857$$

21.3 Estimates of the proportion and number of successes

The estimator of the proportion of successes (items with a specific attribute) in the population, M/N , is obtained by the sample proportion, $\hat{\pi}$. Formally:

$$\hat{\pi} = \frac{y}{n} \quad (21.5)$$

From Equation (21.5), the estimate of the number of successes in the population is:

$$N\hat{\pi} = N \frac{y}{n} \quad (21.6)$$

Thus, if 3 out of a random sample of 12 records were found incomplete, the estimated proportion of incomplete records in the population is $3/12 = 0.25$. If the population contains 65 records, the estimated number of incomplete records in the population is $(0.25)(65) = 16.25$. This number may be rounded down to 16 or reported conservatively as 17.

21.4 Tests of hypotheses

defect For most applications, the attribute of interest is an undesirable property of the item, which we will call a *defect*. With this interpretation, the null and the alternative hypotheses in a hypergeometric experiment are usually set as:

$$H_0: M = M_0 \quad (21.7)$$

$$H_1: M < M_0 \quad (21.8)$$

where M_0 is an acceptable upper limit of defects.

We reject H_0 if $Pr\{y < M_0\} < \alpha$ under the assumption that H_0 is true.

- Typically, M_0 is expressed as a fraction of N . Thus, if no more than 20% defectives can be tolerated in the population, then M_0 is set at $0.20N$. If M_0 is not an integer, round M_0 down to the nearest integer.

Example 21.2 demonstrates the calculation of hypergeometric marginal and cumulative probabilities for hypothesis testing.

Example 21.2. Emergency batteries. A store carries $N = 75$ emergency batteries on the shelf and claims that, under normal use, no more than one-third

($M = 25$ batteries) would fail to hold their charge during the first year. Owing to the cost of the experiment, the store attempts to verify that claim with a sample of size $n = 10$. What is the probability that, if we randomly select 10 batteries, none will fail during the first year of its use? one? two? more than two? How can we test the null hypothesis that the store's claim is correct when $\alpha = 0.05$?

Selected marginal and cumulative probabilities are calculated, using Excel's =HYPGEODIST($y, n, 25, 75$) function, and recorded in Table 21.1 for $n = 10, 11, \text{ and } 12$. (Only the entries for $n = 10$ are needed for this example; the entries for $n = 11$ and $n = 12$ will be needed for Section 21.5.)

Table 21.1. Hypergeometric probabilities for $M = 25, N = 75$

y	n	M	N	Marginal $Pr\{Y = y\}$	Cumulative $Pr\{Y \leq y\}$
0	10	25	75	0.012	0.012
1	10	25	75	0.076	0.088
0	11	25	75	0.008	0.008
1	11	25	75	0.052	0.060
0	12	25	75	0.005	0.005
1	12	25	75	0.036	0.040
2	12	25	75	0.118	0.158

The probabilities of finding $y = 0$ or 1 defective batteries when $n = 10$ are 0.012 and 0.076, respectively. The probability of finding more than one defective battery in the sample is therefore $1 - 0.012 - 0.076 = 0.912$.

To test the store's claim that $M \leq 25$, we set $M_0 = 25$ in Equations (21.7) and (21.8). The claim will then be supported if the null hypothesis is rejected.

We have $M_0 = 25$ and $M_0/N = 0.333$. From Table 21.1, under $H_0: M = 25$, the probability of finding $y = 0$ defective batteries in a sample of size $n = 10$ is $Pr\{0\} = 0.012 < 0.05$. Hence, when $\alpha = 0.05$, if $y = 0$ we reject H_0 and claim statistical evidence that the fraction of defective batteries is less than 0.333.

Suppose that $y = 1$, i.e., one defective battery is found in the sample. From Table 21.1, under H_0 the probability of no more than one defective battery in a sample of size $n = 10$ is $Pr\{0\} + Pr\{1\} = 0.012 + 0.076 = 0.088$. Because $0.088 > 0.05$, we do not reject H_0 , and conclude that we do not have sufficient statistical evidence to claim that the fraction of defective batteries is less than 0.333. Therefore, with $n = 10$, the only result that supports the store's claim is to have no defective batteries in the sample.

This result may not be very reassuring to the store, because if their claim is correct at the margin ($M = 25$), the probability of rejecting H_0 and supporting

their claim is only about 1%. Unless the fraction defective is considerably less than the claimed one-third, H_0 is not likely to be rejected. The probability of rejection is given by the power of the test (see Section 13.4).

The problem here is that the sample is too small to be sensitive to the correctness of H_0 . One solution to this dilemma is, of course, to increase the sample size. This option is considered in Section 21.5.

- Although this example demonstrates the use of a one-sided test, the extension to a two-sided test can be similarly developed. However, it is unlikely that a two-sided test would ever be needed. For NRC applications, we are concerned about being either in compliance or out of compliance, which generally requires a one-sided test.

21.5 Sample size considerations

Chapter 13 considers the relation between the sample size and the power in hypothesis testing when the data came from a normal distribution. In this section, we consider the relation between the sample size and the rejection region. From Table 21.1, we see that, given $N = 75$, $M = 25$, and $\alpha = 0.05$:

- When $n = 10$, we can support $H_1: M < M_0$ if $y = 0$, because
 $Pr\{Y = 0\} = 0.012 < 0.05$, but not if $y = 1$, because
 $Pr\{Y \leq 1\} = 0.088 > 0.05$.
- When $n = 11$, we can support $H_1: M < M_0$ if $y = 0$, but not if $y = 1$, because
 $Pr\{Y \leq 1\} = 0.060 > 0.05$.
- When $n = 12$, we can support $H_1: M < M_0$ if $y = 0$ or if $y = 1$, because
 $Pr\{Y \leq 1\} = 0.040 < 0.05$.

We see that there are various combinations of sample sizes and rejection regions for testing H_0 . We emphasize that the choice of the sample size must be made before the sample is drawn. For example, it is inappropriate to start with a sample of size of $n = 10$ and, if a single defective item is found, to increase the sample to $n = 12$. An explanation why switching from one sample size to another is not permissible once the sampling has started is given, in a different context, in Chapter 24.

As another example, suppose $N = 1000$ and $M/N = 0.05$. Table 21.2 shows selected marginal and cumulative probabilities for $N = 1000$ and $M = 50$.

Table 21.2. Hypergeometric probabilities for $N = 1000$ and $M = 50$

y	n	M	N	Marginal $Pr\{Y = y\}$	Cumulative $Pr\{Y \leq y\}$
0	56	50	1000	0.052	0.052
0	57	50	1000	0.049	0.049
1	57	50	1000	0.157	0.258
0	89	50	1000	0.008	0.008
1	89	50	1000	0.043	0.052
0	90	50	1000	0.008	0.008
1	90	50	1000	0.041	0.049
2	90	50	1000	0.105	0.154

As for Example 21.2, suppose we wish to test $H_0: M = M_0 = 50$ against a left-sided alternative $H_1: M < 50$ when $\alpha = 0.05$. From Table 21.2, we see that:

- When $n = 56$, we cannot support H_1 if $y = 0$ because $Pr\{y = 0\} = 0.052 > 0.05$.
- When $n = 57$, we can support H_1 if $y = 0$ because $Pr\{y = 0\} = 0.049 < 0.05$, but not if $y = 1$, because $Pr\{y \leq 1\} = 0.258 > 0.05$.
- When $n = 89$, we can support H_1 if $y = 0$ because $Pr\{y = 0\} = 0.008 < 0.05$, but not if $y = 1$ because $Pr\{y \leq 1\} = 0.052 > 0.05$.
- When $n = 90$, we can support H_1 if $y = 0$ or $y = 1$ because $Pr\{y \leq 1\} = 0.049 < 0.05$, but not if $y = 2$ because $Pr\{y \leq 2\} = 0.154 > 0.05$.

We see that there are various combinations of sample sizes and rejection regions for testing H_0 . By trial and error, we find that the minimum sample size for testing H_0 is $n = 57$, because $Pr\{y = 0\} > 0.05$ for $n < 57$. If the rejection region for H_0 is increased to $\{y \leq 1\}$, the minimum sample size is 90, because $Pr\{y \leq 1\} > 0.05$ for $n < 90$.

21.6 Normal approximation to the hypergeometric distribution

Probability calculations associated with the hypergeometric distribution are involved and lengthy, especially when cumulative probabilities are required. In some situations, however, the hypergeometric distribution can be approximated by other distributions. In this section, we identify a guideline that allows us to approximate the hypergeometric by a normal distribution. Chapter 22 considers approximation by a binomial distribution.

Hald (1952b), p. 691, states that a normal approximation to the hypergeometric distribution is reasonable if $V(Y) > 9$; that is, from Equation (21.3), if:

$$V[Y] = \frac{nM}{N} \frac{N-M}{N} \frac{N-n}{N-1} > 9 \quad (21.9)$$

If Equation (21.9) holds, then the standard statistic:

$$Z = \frac{Y - E[Y]}{\sqrt{V[Y]}} = \frac{Y - E[Y]}{SD[Y]} \quad (21.10)$$

is distributed approximately as a standardized normal variable. A test of a hypothesis can be made by first substituting the hypothesized value for M/N for $E(Y)$ in the numerator of Equation (21.10) and the calculated value of $SD(Y)$ from Equation (21.4) in the denominator of Z . We then compare the resulting statistic z to a normal quantile $z_{1-\alpha}$. We demonstrate this approximation in Example 21.3.

Example 21.3. Cable installation. A collection of 360 cables is audited for installation compliance. The installer claims that at least 80% of the cable installations are compliant. We examine a sample of size $n = 100$ and find that 12 of the cables in the sample are not in compliance. Do we have enough evidence to challenge the 20% upper limit on noncompliant cable installations claimed by the installer?

In this example, we have $N = 360$ and $n = 100$. The null and alternative hypotheses are $H_0: M = 72$ and $H_1: M < 72$. We use $\alpha = 0.05$.

➤ Note that H_1 is a left-sided hypothesis. This means that H_0 must be rejected to support the installer's claim.

To see if Hald's criterion in Equation (21.9), is met, we calculate $V(Y)$ under the null hypothesis. Using Equation (21.5), we obtain:

$$V[Y] = \frac{(100)(72)}{360} \frac{(360-72)}{360} \frac{(360-100)}{(360-1)} = 11.59$$

Since $11.59 > 9$, we can use the normal approximation to the hypergeometric distribution.

The mean and the standard deviation of Y under H_0 are:

$$E[Y] = \frac{(100)(72)}{360} = 20.00$$

$$SD[Y] = \sqrt{11.59} = 3.40$$

From Equation (21.10), the corresponding Z statistic is:

$$z = \frac{12 - 20.00}{3.40} = -2.35$$

Because $-2.35 < z_{0.05} = -1.645$, we reject H_0 . The statistical evidence supports the claim that the proportion of cables not in compliance is at most 20%.

If we were to use the hypergeometric distribution directly, we would have to calculate and add the probabilities of finding 0, 1, 2, ..., 12 unacceptable cables. Table 21.3 gives these probabilities.

Table 21.3. Hypergeometric probability calculations for Example 21.3

y	0	1	2	3	4	5	6	7
$Pr\{y\}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$Pr\{Y < y\}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
y	8	9	10	11	12	13	14	15
$Pr\{y\}$	0.0001	0.0004	0.0012	0.0030	0.0068	0.0137	0.0249	0.0408
$Pr\{Y < y\}$	0.0002	0.0006	0.0017	0.0047	0.0115	0.0252	0.0501	0.0909

According to Table 21.3, statistical significance is realized for $\alpha = 0.05$ and H_0 is rejected when $y \leq 13$ since $Pr\{Y \leq 13\} = 0.0252$. Using the normal approximation when $y = 13$, we have $z = (13 - 20)/3.40 = -2.05$. From Table T-1 of the appendix, this corresponds to a probability of 0.0202. Continuing the comparison of the exact hypergeometric probabilities with the normal approximation, from Table 21.3 we have $Pr\{Y \leq 14\} = 0.0501$, which is not significant (although it borders on significance). Using the normal approximation for $y = 14$, $z = (14 - 20)/3.40 = -1.76$, which corresponds to a probability of 0.0388 and is statistically significant. When $y = 15$, neither the exact hypergeometric probability nor the normal approximation results in statistical significance.

22

Binomial experiments

22.1 What to look for in Chapter 22

Chapter 22 focuses on the binomial distribution and its application to binomial experiments. The binomial distribution, which is the mathematical foundation of the binomial experiment, is formally presented in Section 8.7, and some of the material from that section will be repeated and reinforced here. As in Section 8.7, this chapter underscores the requirements that qualify experiments as binomial and provides tools for estimation and inferences associated with such experiments.

This chapter shows how to calculate probabilities and make inferences about the structure and composition of the population from which a sample is taken. This chapter will show how to:

- *recognize binomial experiments, §22.2*
- *calculate probabilities associated with the binomial distribution, §22.3*
- *recognize when binomial probabilities may be approximated, §22.3*
- *estimate a population's proportion of attributes, §22.3*
- *test hypotheses about population proportion of attributes, §22.6*

- recognize when a binomial can be approximated by a normal distribution, §22.3
- construct confidence intervals for proportions when the normal approximation is used, §22.6
- construct confidence intervals for proportions when the normal approximation is not used, §22.7

22.2 Prerequisites for a binomial experiment

binomial experiment *Binomial experiments* deal with binary outcomes. However, whereas hypergeometric experiments require sampling without replacement, binomial experiments require sampling with replacement.

The requirements under which an experiment qualifies as a binomial experiment are listed below. Some of these requirements may overlap.

Requirement 1. The experiment is conducted in n trials, all of which are conducted under identical conditions.

- The experiment comprises exactly n trials, and all n trials are conducted under the same set of circumstances. This requirement for identical conditions, however, need not be taken literally. In flipping a coin, for instance, we do not insist that the coin be flipped on the same day or at the same time. For that matter, we may even permit another coin to be flipped. The important principle to remember is that factors that may affect the experiment's results must be the same for all trials in the experiment. As an example, if we wish to assess the probability that pipe joints are welded properly, we must be sure that all the joints in our sample are welded by the same welder, or at least by welders with similar training and certification.

Requirement 2. The i^{th} trial must result in either a success or a failure.

- The response must be binary (dichotomous). Whether we toss a coin or inspect a valve for compliance, the experiment must lead to a success or a failure. Of course, when we define a success, we must have a frame of reference clearly in mind, since success to one person may be a failure to another. To handle the results of a binomial experiment mathematically, we denote the result of the i^{th} trial by Y_i and assign $Y_i = 1$ if the trial yields a success and $Y_i = 0$ otherwise. The advantage of this notation is that the sum $\sum Y_i$ is the number of successes in the sample.

Requirement 3. The probability of success, denoted by π , is constant from trial to trial. As with all probabilities, $0 \leq \pi \leq 1$.

- If, for example, we throw a six-faced die, the probability of rolling a 6 is presumed to be unchanged from one throw to the next. Indeed, a starting model for such an experiment could specify $Pr\{6\} = 1/6$. In contrast, if the quality of the welding improves over time (perhaps owing to experience or to improved equipment), then this assumption of constant probability is violated. Similarly, if the quality of welding deteriorates over time (perhaps owing to fatigue, overconfidence, or mere sloppiness), the binomial experiment is not the correct model to describe the experiment.

Requirement 4. The n trials are independent.

- The system has no memory. For example, if we win eight times in a row in a coin-tossing contest, our chance of winning on the ninth trial should not be different from that of any other toss. Another example is a hiring practice that does not have a quota, and thus, one hiring does not affect who will, or will not, be hired next.

The mathematical treatment of binomial variables involves either or both of two statistics:

$$Y = \sum Y_i, \text{ the number of successes in a sample of } n \text{ trials} \quad (22.1)$$

$$P = Y/n, \text{ the sample proportion of successes in the sample} \quad (22.2)$$

22.3 Binomial probabilities

The binomial probability function gives the probability $Pr\{Y = y\}$ of y successes in n trials when the probability of a success in a single trial is π , $0 \leq \pi \leq 1$. We write this probability function as:

$$\begin{aligned} f(y) &= f(y; n, \pi) \\ &= Pr\{y/n, \pi\} = \binom{n}{y} \pi^y (1-\pi)^{n-y} \\ &= \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}, \quad y = 0, 1, \dots, n \end{aligned} \quad (22.3)$$

The parameters of the binomial distribution are n and π , as indicated in the expression for the probability function. When n and π are obvious from the context, we may find it convenient to write $Pr\{y\}$ instead of $Pr\{Y = y/n, \pi\}$.

Although Equation (22.3) is not especially intimidating or difficult to evaluate, repeated evaluations of binomial probabilities for many different sets of the three arguments y , π , and n can be tedious. Fortunately, we have several “outs.”

First, we have Table T-8, “Selected binomial probabilities” of the appendix. It gives, to four decimal places, the binomial probabilities for combinations of $\pi = 0.01, 0.05, 0.10, 0.25,$ and $0.50,$ $y = 0, 1, \dots, 9$ and $n = 1, 2, \dots, 40.$

Second, we may have access to a variety of published detailed binomial tables. Among them are a pamphlet issued by the Ordnance Corps (1952), a book by Romig (1953), a compilation by the staff of Harvard University’s Computation Laboratory (1955), and tables by Beyer (1974), p. 182.

Third, we may have a hand-held calculator with built-in binomial functions. Then, it is a matter of pressing buttons and reading displays.

Fourth, we may have a spreadsheet program in our desktop computer. Excel’s function `=BINOMDIST($y, n, \pi, 0$)` returns the marginal probability $Pr\{Y = y\}$ and `=BINOMDIST($y, n, \pi, 1$)` returns the cumulative probability $Pr\{Y \geq y\}.$

Fifth, if certain conditions pertaining to the binomial’s parameters are met, we can use the normal approximation, discussed in Section 22.7, to approximate specific probabilities.

Sixth, if certain conditions pertaining to the binomial parameters are met, we can use the Poisson approximation, discussed in Section 23.7, to approximate specific probabilities.

22.4 Examples of binomial experiments

Example 22.1 shows a classical binomial experiment.

Example 22.1. Head count. What is the probability of throwing six heads in a row with a fair coin?

We have $y = 6,$ $n = 6,$ and $\pi = 0.5,$ from which we calculate:

$$Pr\{6\} = \frac{6!}{6!0!} 0.5^0 0.5^{6-0} = 0.0156$$

Table T-8 of the appendix also gives $Pr\{Y = 0, 6, 0.5\} = 0.0156.$

Excel’s function returns `=BINOMDIST(0, 6, 0.5, 0) = 0.0156.`

Because the chance of obtaining six heads in a row is rather small, the fairness of the coin may be suspect.

Example 22.2. Acceptance sampling. A manufacturer advertises that at least 95% of the reed switches produced by his process will function perfectly “out of the box.” To verify his claim, he agrees to a random inspection of

124 switches. The proposed sampling plan states that if the number y of failed switches is no higher than $y_A = 2$, he will be awarded a contract as a vendor for a large order of switches. If $y > 2$, the buyer will look for another vendor.

- In quality assurance, the population from which we draw the sample is referred to as a “lot.”

If the manufacturer is correct in his 95% claim, what is the probability of finding zero, one, or two failed switches in the sample?

The probability that any switch in the sample is defective is $\pi = 0.05$. Setting $n = 124$, we calculate:

$$Pr\{0\} = \frac{124!}{0!124!} (0.05)^0 (0.95)^{124} = 0.0017, \text{ reported here as } 0.002.$$

We continue our calculations and summarize them in Table 22.1.

Table 22.1. Binomial probabilities for $n = 124$ and $\pi = 0.05$

y	Marginal $Pr\{Y = y\}$	Cumulative $Pr\{Y \leq y\}$	Complement $Pr\{Y > y\}$
0	0.002	0.002	0.998
1	0.011	0.013	0.987
2	0.037	0.050	0.950

The probability of the contract award is $Pr\{Y \leq 2\}$. If that probability is small, the contract is awarded. From Table 22.1, the probability of award when $Y \leq 2$ is $1 - 0.050 = 0.950$.

- In the sampling plan just described, the probability of accepting the entire lot is 0.95 when 95% of the items in the lot are “good.” This sampling plan is often referred to as a 95/95 plan. We consider this and similar plans in Chapter 24.

We note that when $n = 124$ and $y = 2$ (122 good items in the sample), the proportion of acceptable switches in the sample is $122/124 = 0.984$, which is better than the required 95%. The margin of 98.4% over 95% is required for assurance that the population (not the sample!) proportion is at least 95%.

Example 22.3. Drug testing. Every employee in a government agency is a candidate for drug testing. In this example, tests are administered once a month, but the day of the month is selected at random. (One problem with this scheme is that once a test is conducted during a given month, employees may be without a deterrence for the rest of that month.) For each of those 12 days, 1/12 (8.33%) of the employees are randomly selected to be tested. The random

sample is selected independently of whether the employee was selected in any previous month. Hence, some employees may be selected more than once in a year, while others may not be selected at all.

Frequently asked questions about random drug testing are: What is the probability that an employee will not be selected during the year? What is the probability that an employee will be selected once, twice, or more than twice during the year?

Drug testing is an excellent opportunity to misuse or interpret the term “sample.” Loosely speaking, we “sample” the population 12 times a year; each time, we draw a “sample” of, say, 100 employees and then we ask every selected employee to give a “sample.”

Technically, the sample given by the selected employee should be called a specimen. The question is then whether the sample size is $n = 12$ or $n = 100$. The answer depends on how we envision the process. If we look at every month as another sample (with $n = 100$), we have a sampling without replacement plan because once an employee is selected, that employee cannot be selected again in the same month.

On the other hand, we can look at the experiment as $n = 12$ trials, where a specific employee may be selected for one or more of these trials. In this context, we have a sampling with replacement plan, which qualifies the experiment as a binomial. This is the scenario in which we answer the question about the probability of being selected y times during the year.

The probability of being selected at a given test is $\pi = 1/12 = 0.0833$. With $n = 100$, we calculate:

$$Pr\{0\} = \frac{12!}{0!12!} (0.0833)^0 (0.9167)^{12} = 0.3521$$

We continue our calculations and summarize them in Table 22.2.

Table 22.2. Binomial probabilities for $n = 12$ and $\pi = 0.0833$

y	Marginal $Pr\{Y = y\}$	Cumulative $Pr\{Y \leq y\}$	Complement $Pr\{Y > y\}$
0	0.3521	0.3521	0.6479
1	0.3840	0.7361	0.2639
2	0.1920	0.9281	0.0719

Using the calculations in Table 22.2, we respond to the questions posed earlier. The probability of an employee not being selected even once during the year is 0.3521, of being selected exactly once during a year of testing is 0.3840, and of being selected exactly twice is 0.1919. The probability of being selected twice or less is 0.9281, and the probability of being selected more than twice is 0.0719.

- The sampling scheme described in Example 22.3 is typically labeled as a “100% sampling rate test.” This term is a common, but unfortunate, because it has the potential to mislead. The label “100% sampling rate” is meant to describe a plan in which the total number of individual drug tests conducted in a specific year is equal to the number of employees in the pool. It does not mean that all employees (i.e., 100%) will be tested.

Example 22.4. Sampling memory chips. Suppose that approximately $\pi = 0.1 = 10\%$ of special computer memory chips are defective. For our computer to run a new batch of software properly, we must install eight of the special memory chips. Our computer dealer carries those chips in a large container, but we can make only one trip to his warehouse. What is the probability that, if we randomly select eight chips, all of them will work?

The probability that all eight chips will function properly (i.e., that zero chips will fail) is given by:

$$Pr\{0\} = \frac{8!}{0!8!} (0.1)^0 (0.9)^8 = 0.4305$$

This probability, slightly larger than 0.43, reflects an uncomfortable “likelihood” that our computer will be operational tonight. Thus, we might decide to buy nine chips to improve our “chances” of winding up with eight working chips. All we have to do is to calculate the probability that the number of defective chips in a sample of $n = 9$ would be either zero or one, because either outcome is acceptable. Table 22.3 gives these calculations.

Table 22.3. Binomial probabilities for $\pi = 0.1$ and varying n

n	y	Marginal $Pr\{Y = y\}$	Cumulative $Pr\{Y \leq y\}$	Complement $Pr\{Y > y\}$
8	0	0.4305	0.4305	0.5695
9	0	0.3874	0.3874	0.6126
9	1	0.3874	0.7748	0.2252
10	0	0.3487	0.3487	0.6513
10	1	0.3874	0.7361	0.2639
10	2	0.1937	0.9298	0.0702
11	0	0.3138	0.3138	0.6862
11	1	0.3835	0.6974	0.3026
11	2	0.2131	0.9104	0.0896
11	3	0.0710	0.9815	0.0185

For eight out of the nine chips to be acceptable, we must have either zero or one bad chip. Table 22.3 gives the associated probability as $Pr\{Y \leq 1\} = 0.7748$, which is a considerable improvement over 0.4705.

If the probability of 0.7748 is still inadequate, then how many chips shall we buy? Even if we buy 50 chips, we have no absolute assurance that at least 8 chips will be acceptable. Hence, we have to set a confidence statement to give us an assurance of at least eight functioning chips. Quite typically, we set that confidence at 95%. Now we reformulate our question to ask how many chips we should buy so that we have at least 95% assurance that at least eight chips will function. From Table 22.3, we see that this assurance may not be achieved unless we buy at least 11 chips (for which we actually have over 98% confidence).

In this example, we set the confidence level at 95%. Because this threshold is arbitrary, we are reminded that the decision concerning the sample size ultimately rests with the user. Finances, convenience, and other considerations will enter into the decision. Here are some guidelines: Recognize and articulate constraints and limitations early in the planning process. Once the goals and decision criteria are decided on, stick to them. Adherence to the chosen goals and decision criteria will keep us out of the trouble we may encounter if we fall back upon *ad hoc*, spur-of-the-moment, changes in our approach.

22.5 Mean and variance of a binomial variable

If we observe Y defective items in a random sample of size n , the sample proportion of defective items is Y/n . This sample proportion is a natural candidate as an estimator of the population proportion π .

Both Y (the sample count of items with the attribute) and Y/n (the sample proportion of items with the attribute) are important statistics. Whereas many statistics books concentrate on the statistic Y , for its mathematical tractability, we prefer to emphasize the statistic Y/n for practical reasons. For notational and mnemonic convenience, we set $P = Y/n$.

Each of Y and P has a mean, a variance, and a standard deviation. As expected, the mean of the proportion P of a sample is the proportion π of successes in the population. Mathematically, we write:

$$E[Y] = n\pi \quad (22.4)$$

$$E[P] = \pi \quad (22.5)$$

$$V[Y] = n\pi(1-\pi) \quad (22.6)$$

$$V[P] = \pi(1-\pi)/n \quad (22.7)$$

$$SD[Y] = \sqrt{n\pi(1-\pi)} \quad (22.8)$$

$$SD[P] = \sqrt{\pi(1-\pi)/n} \quad (22.9)$$

The proof of Equations (22.4) and (22.5) may be found in Hoel (1971), p. 62.

Table 22.4 summarizes the parameters of the binomial probability function.

Table 22.4. Parameters of the binomial probability function

Measure	Sum of successes, Y	Proportion of successes, $P = Y/n$
Mean:	$n\pi$	π
Estimator:	nP	P
Variance:	$n\pi(1-\pi)$	$\pi(1-\pi)/n$
Estimator:	$nP(1-P)$	$P(1-P)/n$
Standard deviation:	$\sqrt{n\pi(1-\pi)}$	$\sqrt{\pi(1-\pi)/n}$
Estimator:	$\sqrt{nP(1-P)}$	$\sqrt{P(1-P)/n}$

22.6 Normal approximation to the binomial

Although Y always is an integer in binomial experiments, certain conditions allow us to approximate the distribution of Y and, in parallel, the distribution of $P = Y/n$, by the ubiquitous normal distribution. This approximation simplifies the term-by-term calculations inherent in the binomial by employing the easy-to-use standard normal table (such as Table T-1 of the appendix) and allowing solutions to many problems without the need for an extensive enumeration exercise. The conditions for using the normal approximation are stated by writers such as Dixon and Massey (1983), p. 220, as:

$$\text{Both } n\pi \geq 5 \text{ and } n(1 - \pi) \geq 5 \quad (22.10)$$

If the binomial parameter π is not known, we can use $P = Y/n$ to approximate π , yielding the following modified equivalent conditions:

$$\text{both } nP \geq 5 \text{ and } n(1 - P) \geq 5 \quad (22.11)$$

$$\text{both } n(Y/n) \geq 5 \text{ and } n(1 - Y/n) \geq 5 \quad (22.12)$$

$$\text{both } Y \geq 5 \text{ and } (n - Y) \geq 5 \quad (22.13)$$

When these conditions are satisfied, we can call on all the procedures involved in normal estimation and hypothesis testing processes. Thus, we may test whether π is equal to a given value (using either a one- or a two-sided test), or we can produce a confidence interval (either one- or two-sided interval) for π .

To test the null hypothesis $H_0: \pi = \pi_0$, the test statistic is:

$$Z = \frac{Y/n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \quad (22.14)$$

We reject H_0 if:

$$H_1: \pi \neq \pi_0 \text{ and } |z| > z_{\alpha/2} \quad (22.15)$$

$$H_1: \pi < \pi_0 \text{ and } z < -z_{1-\alpha} \quad (22.16)$$

$$H_1: \pi > \pi_0 \text{ and } z > z_{1-\alpha} \quad (22.17)$$

Examples 22.5 and 22.6 illustrates the test.

Example 22.5. Water impurity. From a large area near a plant, 63 aquifers were randomly selected and examined for impurities. Excessive contaminants were found in 16 aquifers. Is the sample proportion significantly higher than the prevalent 10% in that county?

We have $n = 63$, $y = 16$, $p = y/n = 16/63 = 0.254$. The setup for the test is $H_0: \pi = 0.10$, $H_1: \pi > 0.10$, and $\alpha = 0.05$.

First note that, under H_0 , Equation (22.13) is satisfied as $np = 16 > 5$ and $n(1-p) = 144 > 5$, so that the normal approximation to the binomial is appropriate. We proceed by constructing a standard normalized test statistic and reject $H_0: \pi = 0.1$ if the calculated statistic is larger than $z_{0.95} = 1.645$. From Equation (22.14), we have:

$$z = \frac{16/63 - 0.1}{\sqrt{\frac{0.1(1-0.1)}{63}}} = 4.07$$

Because $4.07 > 1.645$, we reject H_0 , claiming statistical evidence that the level of contamination is larger than 10%.

Example 22.6. Mask filters. Out of 100 masks tested in a random sample, 1 failed to meet a standard criterion. Since this may be a serious issue, we wish a proof that no more than $\pi = 5\%$ of the entire lot is unacceptable. Since, under H_0 , $\pi n = (0.05)(100) = 5$, we use the normal approximation. We set the hypothesis in such a way that the test has to show that the percent of unacceptable masks is significantly less than 5%. We have:

$H_0: \pi = 0.05$, $H_1: \pi < 0.05$, $\alpha = 0.05$

$$z = \frac{1/100 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{100}}} = -1.84$$

The critical point for this test is $z_{0.05} = -1.645$. Because $-1.84 < -1.645$, we reject the null hypothesis in favor of a statement that the population proportion of unacceptable filters is below 5%.

22.7 Confidence interval for a proportion: **Normal approximation applies**

When the normal approximation applies, that is $Y \geq 5$ and $(n - Y) \geq 5$, the construction of the confidence interval is similar to the procedure used for the population mean in Chapter 9.

A two-sided confidence interval about π is constructed using the sample proportion $P = Y/n$:

$$\text{two-sided: } P \pm z_{1-\alpha/2} \sqrt{\frac{P(1-P)}{n}} \quad (22.18)$$

For a one-sided confidence interval, use, as needed, either one of the following:

$$\text{upper-sided: } P + z_{1-\alpha} \sqrt{\frac{P(1-P)}{n}} \quad (22.19)$$

$$\text{lower-sided: } P - z_{\alpha} \sqrt{\frac{P(1-P)}{n}} \quad (22.20)$$

Example 22.7. Computer updates. Of 100 computers selected at random from a large population, 7 are found to be outdated. What is the 95% upper confidence interval (UCL) on the fraction of outdated computers?

$$\text{UCL} = \frac{7}{100} + 1.645 \sqrt{\frac{7/100 \cdot 93/100}{100}} = 0.07 + 0.042 = 0.112$$

This sets the 95% upper limit for the fraction of cables not meeting standard specifications at slightly more than 11%.

22.8 Confidence interval for a proportion: Normal approximation does not apply

If either $Y < 5$ or $n - Y < 5$, the normal approximation to the binomial cannot be used to construct a confidence interval about the binomial proportion π . However, other approaches are available.

Given a sample of size n with Y successes, we start by constructing a two-sided $100(1 - \alpha)\%$ confidence interval on the binomial parameter π . Denote the interval by (π_L, π_U) . Let y be the observed number of successes, where $0 < y < n$. Then π_U is the value of π for which:

$$\sum_{i=0}^y \frac{n!}{i!(n-i)!} \pi^i (1-\pi)^{n-i} = \alpha/2 \quad (22.21)$$

and π_L is the value of π for which:

$$\sum_{i=y}^n \frac{n!}{i!(n-i)!} \pi^i (1-\pi)^{n-i} = \alpha/2 \quad (22.22)$$

If $0 < y < n$ and we wish to construct a one-sided confidence interval about π , we replace $\alpha/2$ in Equation (22.21) or (22.22) by α for upper and lower bounds, respectively. If $y = 0$, we cannot rule out a value of 0 for π and therefore only a one-sided upper confidence interval makes sense. Similarly, if $y = n$, we cannot rule out a value of 1 for π and therefore only a one-sided lower confidence interval makes sense. These two special cases are discussed in Section 22.9.

Equations (22.21) and (22.22) can be solved either by interpolating in a table of cumulative binomial probabilities or by iteratively solving a polynomial equation in π . Alternatively, as shown by Bowker and Lieberman (1972), p. 466, both π_U and π_L can be approximated by quantiles of the *F-distribution*. However, interpolation in a table of the *F-distribution* (Table T-4 of the Appendix) may be required for acceptable approximations to π_U and π_L . An easy alternative to table interpolation is to use a spreadsheet function, such as Excel's =FINV, to calculate the required *F* values

Using Bowker and Lieberman's approach, we have:

$$\pi_U = \frac{y+1 f_{1-\alpha/2}(2(y+1), 2(n-y))}{(n-y) + y+1 f_{1-\alpha/2}(2(y+1), 2(n-y))} \quad \begin{array}{l} \text{(Two-sided upper} \\ \text{confidence limit for } \pi) \end{array} \quad (22.23)$$

$$\pi_L = \frac{y}{y + n-y+1 f_{1-\alpha/2}(2(n-y+1), 2y)} \quad \begin{array}{l} \text{(Two-sided lower} \\ \text{confidence limit for } \pi) \end{array} \quad (22.24)$$

where $f_q(v_1, v_2)$ denotes the q^{th} quantile of the *F-distribution* (Table T-5 of the appendix).

In Excel, we use =FINV($\alpha/2$, $2(y+1)$, $2(n-y)$) for the required *f* values in Equation (22.23) and =FINV($\alpha/2$, $2(n-y+1)$, $2y$) in Equation (22.24).

If a one-sided confidence interval is needed, then replace $1 - \alpha/2$ by $1 - \alpha$ in both Equations (2.23) and (22.24) to yield:

$$\pi_U = \frac{y+1 f_{1-\alpha}(2(y+1), 2(n-y))}{(n-y) + y+1 f_{1-\alpha}(2(y+1), 2(n-y))} \quad \begin{array}{l} \text{(One-sided upper} \\ \text{confidence limit for } \pi) \end{array} \quad (22.25)$$

$$\pi_L = \frac{y}{y + n - y + 1} \frac{1}{f_{1-\alpha}(2(n-y+1), 2y)} \quad \begin{array}{l} \text{(One-sided lower} \\ \text{confidence limit for } \pi) \end{array} \quad (22.26)$$

In Excel, we use =FINV(α , 2($y+1$), 2($n-y$)) for the required f values in Equation (22.25) and =FINV(α , 2($n-y+1$), 2 y) in Equation (22.26).

Example 22.8. Two-sided confidence interval for proportion.

Bowker and Lieberman (1972, p. 467) illustrate their approach with the following example. Suppose that 4 defects are observed in sample of 25. The sample proportion is $p = 0.16$. We wish to construct a two-sided 95% confidence interval for π with $\alpha = 0.05$, $n = 25$, $y = 4$. Using Equations (22.23) and (22.24), we find $f_{0.975}(2(4+1), 2(25-4)) = f_{0.975}(10, 42) = 2.37$ and $f_{0.975}(2(25-4+1), 2(4)) = f_{0.975}(44, 8) = 3.82$. We then have:

$$\pi_U = \frac{5(2.37)}{21+5 \cdot 2.37} = 0.361$$

$$\pi_L = \frac{4}{4+22 \cdot 3.82} = 0.045$$

Thus, we have 95% confidence that the true population proportion π is between 0.045 and 0.361.

Example 22.9. One-sided confidence interval for proportion.

For most applications, a more realistic objective for Example 22.8 would be to construct a one-sided upper confidence limit for the proportion of defects in the population. From Equation (22.25), we find $f_{0.95}(10, 42) = 2.06$ and then calculate the 95% upper confidence limit for π as:

$$\pi_U = \frac{5(2.06)}{21+5 \cdot 2.06} = 0.329$$

22.9 Confidence interval for a sample with no defects

It often happens that $y = 0$ in a binomial sample of size n , where π is the population proportion of defects. First, we note that the normal approximation does not apply because Equation (22.13) is violated. With no defects in the sample, a lower bound of 0 for π cannot be ruled out. Consequently, we look for an upper confidence limit for π .

From Equation (22.21) with $y = 0$ and $\alpha/2$ replaced by α , the upper confidence limit π_U is the value of π such that:

$$\frac{n!}{0!(n-0)!} \pi^0 (1-\pi)^{n-0} = \alpha \quad (22.27)$$

Since $\frac{n!}{0!n!} = 1$ and $\pi^0 = 1$, Equation (2.27) is reduced to:

$$(1-\pi)^n = \alpha \quad (22.28)$$

To find the value of π satisfying Equation (22.28), we take the logarithm of both sides and then exponentiate to get:

$$n \ln(1-\pi) = \ln(\alpha)$$

$$\ln(1-\pi) = \ln(\alpha) / n$$

$$1 - \pi = e^{[\ln(\alpha)/n]}$$

It follows that:

$$\pi_U = 1 - e^{[\ln(\alpha)/n]} \quad (\text{one-sided upper } (1 - \alpha) \text{ confidence limit}) \quad (22.29)$$

Example 22.10. Contaminated wells. We wish to establish a limit on the proportion of wells in a population that are contaminated. We draw a random sample of 25 wells and find none contaminated. From Equation (22.29), an upper 95% confidence limit is:

$$\pi_U = 1 - e^{\ln(0.05)/25} = 1 - e^{-2.996/25} = 1 - e^{-0.120} = 1 - 0.887 = 0.113$$

Thus, we are 95% confident that no more than 11.3% of all the wells are contaminated.

Suppose our goal is to demonstrate that only a few percent of wells, e.g., 3%, are contaminated. If we are satisfied with only 90% confidence, then the UCL is 8.8%. With 85% confidence, the upper limit decreases to 7.3%, and with 80% confidence, to 6.2%. The UCL is 3% with only 53% confidence. Hence, it may not be possible to achieve our goal with any reasonable confidence level. The only way to significantly improve these results is to increase the sample size. Of course, the only way to be certain that none of the wells is contaminated is to test the entire population.

22.10 Sample size for political polling

In Example 22.10, we saw that, with a fixed sample size, the only way to decrease an upper confidence limit was to decrease the confidence level and,

furthermore, that a desired upper limit may not be achievable with any reasonable confidence level. In this section, we show how increasing the sample size allows us to achieve a target goal with pre-specified confidence.

We illustrate this with a polling example, a widely used application of binomial confidence intervals. Suppose we need to determine the sample size required to predict a forthcoming national election with a specified margin of error. We assume that two candidates are running neck-and-neck, so that the probability of either candidate being elected is approximately 0.5.

We make the following assumptions:

- There are only two candidates, A and B.
- The persons polled are selected at random.
- Each person polled will unequivocally favor either A or B.
- Each person polled will vote for the same candidate on Election Day.

Given a sample of size n , let Y be the number of people who support Candidate A. The ratio $P = Y/n$ is an estimate of π , the proportion of the electorate that will vote for Candidate A on Election Day. For national pre-election polls, the margin of error is typically set at $0.03 = 3\%$. This means that the end points of a two-sided 95% confidence interval about π will differ from π by no more than 3 percentage points or 0.03.

Because n is large (typically about a thousand) and $y \approx n/2$, Equation (22.13) holds and we can use the normal approximation to the binomial. From Equation (22.18), a two-sided 95% confidence interval for π is:

$$P \pm z_{0.975} \sqrt{\frac{P(1-P)}{n}} \quad (22.30)$$

Because the candidates are running neck-and-neck, we set $P = 0.50$. Setting $z_{0.975} = 1.960$, we can satisfy the 3% margin of error requirement by equating half-width of the interval in Equation (22.30) to 0.03 to get:

$$1.960 \sqrt{\frac{(0.50)(0.50)}{n}} = 0.03 \quad (22.31)$$

Solving Equation (22.31) for n yields $n = 1067.1$, which is rounded up to 1068, and often rounded up further to 1100. Many pre-election polling reports include a statement about a 3% “margin of error.” Some also state that the results are based on a sample size of $n = 1100$, and some even explain that the margin of error corresponds to a 95% confidence level.

When the cost of sampling is high, we may reduce the sample size by accepting a larger margin of error. For example, if we use a margin of error of 0.05,

Equation (22.31) yields a sample size of 385. If we are willing to use a lower confidence level, say 90% (for which $z_{0.95} = 1.645$), we can reduce the sample size to 271.

22.11 Binomial approximation to the hypergeometric distribution

sampling ratio

Under certain conditions that are often easy to meet, binomial probabilities can be used to approximate hypergeometric probabilities. Brownlee (1965, p. 167) states: “A rough criterion for the validity of the approximation is that the *sampling ratio* $n/N < 0.1$; i.e., for the sample size to be less than 10 per cent of the population size.” Furthermore, the smaller the sampling ratio n/N , the better the approximation. The main advantages of the binomial over the hypergeometric distribution are that binomial probabilities are easier to calculate and there are more tables available for this task.

Table 22.5 compares hypergeometric and binomial probabilities for a sample of size $n = 10$ from a population of size $N = 100$ with $M = 25$. Because the probability of an item in the population having the attribute of interest is $M/N = 25/100 = 0.25$, we set $\pi = 0.25$ for the binomial distribution.

Table 22.5. Hypergeometric and binomial probabilities compared

Successes y in sample of $n = 100$	Hypergeometric distribution: $Pr\{Y = y N = 100,$ $M = 25, n = 10\}$	Binomial distribution $Pr\{Y = y n = 10,$ $\pi = 25/100\}$	Difference
0	0.048	0.056	-0.008
1	0.181	0.188	-0.006
2	0.292	0.282	0.011
3	0.264	0.250	0.013
4	0.147	0.146	0.001
5	0.053	0.058	-0.005

For this example, the two sets of probabilities differ by no more than 0.01. Note that the sampling ratio for this example is 0.1, which is the minimal requirement for using the binomial approximation. In addition to improving as the sampling ratio decreases, the approximation also improves as the sample size increases.

23

Poisson experiments

23.1 What to look for in Chapter 23

The Poisson distribution⁹ was formally introduced in Section 8.10. This chapter elaborates on the distribution and shows its applications.

The Poisson distribution is a discrete distribution that is often used as a model for calculating probabilities of rare events associated with time and space. For example, with the help of the Poisson distribution function, we will be able to calculate the probabilities of zero, one, two, or more failures of a component with a known constant failure rate, provided that we can satisfy certain prerequisites. These prerequisites, some of which are shared with the prerequisites for the binomial experiments discussed in Chapter 22, are detailed in Section 23.2. This chapter shows us how to:

- *recognize Poisson experiments, §23.2*
- *calculate the probabilities associated with rare events, §23.3*
- *test hypothesis about a Poisson parameter, §23.5*

⁹ After the French mathematician Siméon Denis Poisson (1781–1840).

- *construct a confidence interval for a Poisson parameter, §23.5*
- *approximate the binomial by the Poisson distribution, §23.6*
- *approximate the Poisson by the normal distribution, §23.7*

23.2 Prerequisites for a Poisson experiment

The Poisson distribution is often used as a model for calculating probabilities of the occurrence of rare events in a fixed time interval or in a fixed interval, area, or volume in space. Indeed, every experiment that counts events is a potential candidate for modeling by the Poisson distribution. Some examples of data that can be modeled by the Poisson distribution are counts of:

- trucks overturned on the Beltway during a Winter month
 - babies born in a cab on the way to the hospital during a specific week
 - pumps failing to start at the push of the start button
 - lightning strikes within a one-kilometer circle around a specific power station during a specific month
 - deer hit by motorists on Interstate 270 on a given day
 - alpha particles emitted from a specific source in a 15-second time interval
 - kinks in 1,000 yards of stretched copper wire
 - flaws in 100 continuous yards of fabric
- In the following discussion, we consider Poisson experiments occurring in time. The discussion, however, is equally applicable to experiments occurring in space (length, area, volume).

The requirements, or assumptions, for the application of the Poisson distribution are listed below. We consider the probabilities of events occurring in small subintervals of a given period of time,

- See Requirement 2 for a definition of what is meant by “small” in this context.

*scale
parameter*

Requirement 1. The probability of observing one event is approximately proportional to the length of Δt . That is, $Pr\{1 \text{ event in } \Delta t\}$ is approximately equal to $\theta\Delta t$ for some constant $\theta > 0$. Here, θ is the *scale parameter* of the Poisson experiment.

Requirement 2. The probability of observing more than one event in the subinterval Δt is negligible.

- A subinterval that satisfies Requirement 2 is considered small.

Requirement 3. The occurrence of an event in any small subinterval is independent of the occurrence of an event in any other nonoverlapping subinterval.

- The system has no memory. What took place yesterday does not affect what will happen today.

Requirement 4. The probability approximations in Requirements 1 and 2 improve as Δt approaches zero.

- This assumption allows the rigorous mathematical derivation of the Poisson distribution from the other assumptions.

Because any interval can be thought of as a collection of adjacent nonoverlapping subintervals, it is a logical consequence of the assumptions that the probability of y events occurring in an interval of any length depends only on y and the length of the interval.

- The probability of observing y events in a one-hour interval starting now is the same as for any other one-hour interval starting anytime after an hour from now.

23.3 Poisson probabilities

Given a time interval of length t , let the random variable Y be the total number of events occurring in a Poisson experiment with scale parameter θ . The probability that Y equals y is given by:

$$Pr\{Y = y | \theta t\} = \frac{e^{-\theta t} (\theta t)^y}{y!}, \quad y = 0, 1, 2, \dots \quad (23.1)$$

- The Poisson distribution counts the number of events in a time interval of fixed length. This is in contrast to the binomial distribution, which counts the number of events (successes) in a fixed number of trials.
- If θ is small, from Equation (23.1) with $t = 1$, we have $Pr\{Y = 1\} = \theta e^{-\theta} \cong \theta$. In other words, for small θ , the probability of exactly one event in a unit time interval is approximately equal to θ .

The mean of the random variable Y is (Mood et al. (1974), p. 93):

$$E[Y] = \theta t \quad (23.2)$$

The variance of Y , interestingly, is equal to the mean (ibid), so we write:

$$V[Y] = \theta t \quad (23.3)$$

and the standard deviation of Y is:

$$SD[Y] = \sqrt{\theta t} \quad (23.4)$$

Poisson parameter From Equation (23.1), we see that the Poisson distribution depends on a single **Poisson parameter**, denoted by λ . Setting $\lambda = \theta t$, we have:

$$Pr\{Y = y | \lambda\} = \frac{e^{-\lambda}(\lambda)^y}{y!}, \quad y = 0, 1, 2, \dots \quad (23.5)$$

$$E[P] = \lambda \quad (23.6)$$

$$V[Y] = \lambda \quad (23.7)$$

and the standard deviation of Y is:

$$SD[Y] = \sqrt{\lambda} \quad (23.8)$$

The probabilities associated with various values of λ and y are easily calculated on a hand-held scientific calculator, provided you don't overflow its memory. Alternatively, you can use a source like Table T-9 "Selected Poisson probabilities," of the appendix, which provides probabilities selected values of λ and y .

We can also obtain the marginal probability $Pr\{Y = y | \lambda\}$ using Excel's function =POISSON($y, \lambda, 0$), and the cumulative probability $Pr\{Y \leq y | \lambda\}$ using Excel's function =POISSON($y, \lambda, 1$). As examples, we may wish to verify that:

$$=POISSON(0, 2.5, 0) = 0.082$$

$$=POISSON(1, 2.5, 0) = 0.205$$

$$=POISSON(2, 2.5, 0) = 0.256$$

$$=POISSON(2, 2.5, 1) = 0.544$$

23.4 Applications

Two examples of the application of Poisson probabilities are shown below.

Example 23.1. Errors in message transmission. A status report of plant operation is continuously updated and transmitted to the plant manager electronically. Because the rate of transmission errors is low and constant and because they are independent, the number of transmission errors has a Poisson distribution. From previous experience, the probability of transmission error in any one-hour segment is about one percent. Therefore, we assume that the scale parameter $\theta = 0.01$. What is the probability that during a 10-hour period no errors will be found? Exactly one error? Exactly two errors? More than two errors?

Setting $t = 10$ and $\theta = 0.01$, we have $\lambda = (10)(0.01) = 0.10$. The marginal probabilities of exactly zero, one, or two errors can be calculated from Equation (23.1) or obtained from Table T-9 of the Appendix $\lambda = 0.10$ and $y = 0, 1, \text{ and } 2$. The probability of more than two errors is the complement of at most two errors. We have:

$$Pr\{0\} = Pr\{0 \mid 0.10\} = (e^{-0.1})(0.1)^0/0! = (0.9048)(0.1)^0/0! = 0.9048$$

$$Pr\{1\} = (0.9048)(0.1)^1/1! = 0.0905$$

$$Pr\{2\} = (0.9048)(0.1)^2/2! = 0.0045$$

$$\begin{aligned} Pr\{\text{more than } 2\} &= 1 - Pr\{0\} - Pr\{1\} - Pr\{2\} \\ &= 1 - 0.9048 - 0.0905 - 0.0045 = 0.0002 \end{aligned}$$

➤ The calculated probabilities apply to any number of non-overlapping intervals that total to 10 hours.

Example 23.2. Elevator failure. The probability of an elevator failure is certified to be 0.014 in an arbitrary 24-hour period. What is the average number of elevator failures in a 5-day workweek? What is the probability of no failure in 5 consecutive days? In 5 consecutive Mondays?

We set the scale parameter $\theta = 0.014$. Setting $t = 5$, the average number of failures over five consecutive days is $\lambda = (5)(0.014) = 0.70$. From Equation (23.1), the probability of no failure is calculated as:

$$Pr\{0\} = (e^{-0.70})(0.70)^0/0! = (0.4966)(0.1)^0/0! = 0.4966.$$

Thus there is nearly a 50% chance of at least one elevator failure during the week.

The probability of no failure on five consecutive Mondays is 0.4966, the same as in any five non-consecutive days, or any collection of non-overlapping time intervals that sum up to five days.

23.5 Parameter testing and confidence intervals

As for other distributions, there are several options for testing whether the Poisson parameter is equal to a given value, λ_0 . These are:

$$H_0: \lambda = \lambda_0, H_1: \lambda \neq \lambda_0 \quad (\text{two-sided test}) \quad (23.9)$$

$$H_0: \lambda = \lambda_0, H_1: \lambda > \lambda_0 \quad (\text{one-sided test}) \quad (23.10)$$

$$H_0: \lambda = \lambda_0, H_1: \lambda < \lambda_0 \quad (\text{one-sided test}) \quad (23.11)$$

➤ Most applications of interest consider only one-sided alternatives, since we are usually concerned about λ being too large or too small.

Because of the equivalence between hypothesis testing and confidence interval, we will test a null hypothesis about λ by constructing a confidence interval for λ . If the event count falls outside the confidence limits, we will reject H_0 and conclude that the statistical evidence is not consistent with the null hypothesis.

As shown by Brownlee (1965), p 172, we can form a confidence interval about λ through the chi-square distribution, pertinent values of which are given in Table T-2 of the appendix or obtained from Excel's =CHIINV function.

To construct a $100(1 - \alpha)\%$ confidence interval about λ , we first determine Y , the count of events in total time t . We then calculate the following degrees of freedom:

$$v_L = 2Y \quad (23.12)$$

$$v_U = 2(Y + 1) \quad (23.13)$$

For a two-sided confidence interval, the lower and upper $100(1 - \alpha)\%$ confidence limits about λ are:

$$\lambda_L = \frac{1}{2} \chi_{\alpha/2}^2 \quad v_L \quad (23.14)$$

$$\lambda_U = \frac{1}{2} \chi_{1-\alpha/2}^2 \quad v_U \quad (23.15)$$

➤ If $y = 0$, set the lower limit of the confidence interval at $\lambda_L = 0$.

When $y > 0$, a one-sided $100(1 - \alpha)\%$ lower confidence limit on λ is:

$$\lambda_L = \frac{1}{2} \chi_{\alpha}^2 \quad \nu_L \quad (23.16)$$

If $y = 0$, set $\lambda_L = 0$.

When $y > 0$, a one-sided $100(1 - \alpha)\%$ upper confidence limit on λ is:

$$\lambda_U = \frac{1}{2} \chi_{1-\alpha}^2 \quad \nu_U \quad (23.17)$$

Example 23.3. Automobile accidents. Suppose we wish to test whether the average number of accidents in a given intersection is no larger than five a year, when the record shows $y = 7$ accidents for the last year. Let $\alpha = 0.05$. From Equations (23.13) and (23.17), the upper 95% confidence limit is:

$$\nu_U = 2(y + 1) = (2)(8) = 16$$

$$\lambda_U = \frac{1}{2} \chi_{0.95}^2 \quad 16 = \frac{1}{2} \quad 26.3 = 13.15$$

Because $7 < 13.15$, we do not reject the null hypothesis and conclude that the statistical evidence does not contradict the claim that the average annual number of accidents in the given intersection is no larger than five.

An alternative to using Equations (23.14) to (23.17) to construct confidence limits is to use Table T-10 of the appendix for selected values of y from 0 to 50 (denoted by p in the table) and for several choices of α . However, for one-sided confidence intervals, it is necessary to use the confidence level corresponding to $\alpha/2$ in the table. Thus, for Example 23.3, we find $\lambda_U = 13.15$ in the row for $p = 7$ and the column for $\alpha/2 = 0.05$.

Example 23.4. False alarms. The number of false alarms in a security zone during a 1-month period is treated as a Poisson variable. Suppose that, in a specific month, two false alarms were observed. Based on these data, we seek a one-sided, upper 95% limit for the average number of monthly false alarms.

Using Table T-10 with $y = 2$ and $1 - \alpha = 0.95$, we find the upper limit on the expected number of false alarms to be 7.22.

23.6 Approximations to the binomial distribution

A second major application of the Poisson distribution is as an approximation to the binomial distribution. Recall from Chapter 22 (Equation (22.10)) that the normal distribution provides a good approximation to the binomial distribution if $n\pi \geq 5$. If $n\pi < 5$, it turns out that the Poisson distribution with $\lambda = n\pi$ provides a good approximation to the binomial. A proof may be found in Hoel (1971), p. 64.

The Poisson approximation to the binomial is illustrated with two examples.

Example 23.5. Failures of coating adhesions. The adhesion of urethane coating to concrete surfaces may be tested by a "scratch test" which is applied to a block of concrete with one square foot of surface area. The manufacturer of a coating claims a 99% adhesion rate. If 12 one-square-foot blocks of concrete are selected at random, what is the probability that at least 11 out of the 12 will pass the scratch test? This is equivalent to asking what is the probability that none or one of the scratch tests will fail.

For the binomial calculations, we have $n = 12$ and $\pi = Pr\{\text{scratch test failure}\} = 0.01$. From Equation (22.3), the probabilities of 0 and 1 failure are:

$$Pr\{0|12,0.01\} = \frac{12!}{0!12!} 0.01^0 0.99^{12} = 0.8864$$

$$Pr\{1|12,0.01\} = \frac{12!}{1!11!} 0.01^1 0.99^{11} = 0.1074$$

Therefore, the probability that at least 11 out of 12 blocks will pass the scratch test is $0.8864 + 0.1074 = 0.9938$.

To apply the Poisson approximation, we first check whether $n\pi < 5$. We have $n\pi = 0.12$, so we can use a Poisson approximation. Setting $\lambda = 0.12$, we may obtain the required Poisson probabilities from Table T-9 or use Equation (23.1) to get:

$$Pr\{0\} = (e^{-0.12})(0.12^0)/0! = 0.8869$$

$$Pr\{1\} = (e^{-0.12})(0.12^1)/1! = 0.1064$$

$$Pr\{\text{at least 11 acceptable blocks}\} = Pr\{0\} + Pr\{1\} = 0.8869 + 0.1064 = 0.9933$$

Comparing these results with the exact binomial probabilities, we see that the Poisson approximation is excellent, differing with the binomial by 0.001 or less.

Example 23.6. Misprint Errors. A large official publication has misprint errors on 5% of its pages. If a random sample of 40 pages is examined, what is the probability that no pages in error will be found? Exactly one page with at least one error? Exactly two pages with at least one error each?

We first find the exact probabilities using the binomial distribution and then use the Poisson approximation. Although the probabilities can be easily calculated with a hand-held calculator, we use Table T-8 ($n = 40$, $\pi = 0.05$) for the binomial probabilities and Table T-9 ($\lambda = n\pi = (40)(0.05) = 2.0$) for the Poisson probabilities. These probabilities are summarized in Table 23.1. In the ranges of the values of interest, the two distributions agree to within 0.01 of each other.

Table 23.1. Binomial and Poisson probabilities compared

Event probabilities	Binomial probabilities ($n = 40$, $\pi = 0.05$)	Poisson probabilities ($\lambda = n\pi = 2.0$)
$Pr\{0\}$	0.1285	0.1353
$Pr\{1\}$	0.2706	0.2707
$Pr\{2\}$	0.2777	0.2707

- The close agreement between the binomial and the Poisson distributions is not surprising because, historically, the Poisson distribution was developed as a limiting distribution which approximates the binomial when n is large and π is small, and $n\pi$ is constant. The approximation improves as n increases or π decreases.

23.7 The normal approximation to the Poisson distribution

We end this chapter by discussing another approximation: When λ is large, the Poisson distribution can be approximated by the normal distribution. How large λ needs to be for the approximation to be useful depends on the application. For most practical cases, $\lambda = 10$ suffices and the larger λ is, the better the approximation. Because the mean and variance of a Poisson distribution are equal, a Poisson distribution Y with mean λ is approximated by the normal $N(\lambda, \lambda)$ distribution. Therefore, the standard statistic:

$$Z = \frac{Y - \lambda}{\sqrt{\lambda}} \quad (23.18)$$

is distributed approximately as a $N(0, 1)$ variable.

Figure 23.1 shows the Poisson distribution for $\lambda = 2, 5,$ and $10,$ with a smooth curve connecting the points of the Poisson. Note that when $\lambda = 10$ the Poisson distribution is visually indistinguishable from a normal distribution.

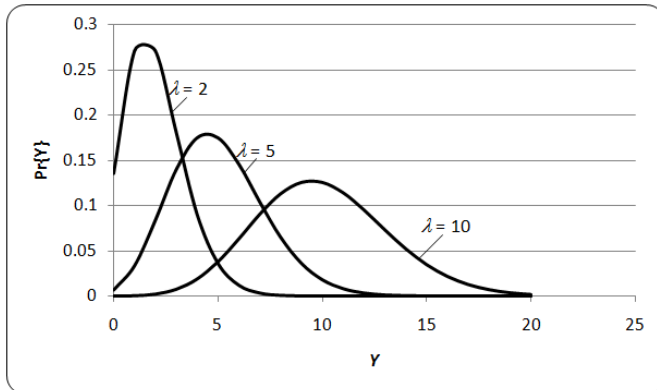


Figure 23.1. Poisson distribution for $\lambda = 2, 5,$ and 10

Example 23.7. Power interruption. A large warehouse experiences an average of 10 power interruptions a month. What is the probability that we observe no more than 16 interruptions during any month?

We will first use the normal approximation and then check it against the exact answer,

For the normal approximation, we set $\lambda = 10$ in Equation (23.18) to obtain:

$$z = \frac{16 - 10}{\sqrt{10}} = 1.897$$

From Table T-1 for the standard normal distribution, $Pr\{Z \leq 1.897\} = 0.971,$ Hence the probability of no more than 16 interruptions during the month is about 0.97.

To check on the normal approximation, we need the Poisson cumulative distribution for $\lambda = 10$ and $y = 16.$ We use Excel's function =POISSON(16,10,1), which returns the value of $Pr\{Y \leq 16 | 10\} = 0.973.$ This is in close agreement with the normal approximation of 0.971.

24

Quality assurance

24.1 What to look for in Chapter 24

Chapter 24 focuses on those concepts that result in procedures designed and constructed to further the pursuit and the achievement of quality. Those procedures may alert a manufacturer that a particular product may not meet its specifications, or they may call a manager's attention to a fault in a service system. Whatever the specific application, the concepts—and the procedures that evolve from them—are collected under the general rubric “quality assurance.” In connection with these quality-driven procedures, we will encounter a number of concepts and specialized terms:

- *quality control*, §24.2
- *process control* §24.3
- *control charts for means* §24.4
- *run test for testing randomness*, §24.5
- *control charts for variability*, §24.6
- *control charts for attributes*, §24.7
- *acceptance sampling*, §24.8.

24.2 The concept of quality assurance

quality assurance

In its broadest terms, *quality assurance* encompasses all the things that an organization does to ensure that its products and services provide value and the expected performance. Quality assurance efforts are usually associated with products, from sampling a day's production of automobiles for fuel efficiency to proofreading responses to customer complaints to funding research into advanced technology. But the basics of quality assurance apply equally to the treatment of employees in an organization or emergency plumbing repairs in residences.

U.S. Code of Federal Regulations, Title 10 (Energy), Part 50, Appendix B (2009), states:

...“quality assurance” comprises all those planned and systematic actions necessary to provide adequate confidence that a structure, system, or component will perform satisfactorily in service. Quality assurance includes quality control which comprises those quality assurance actions related to the physical characteristics of a material, structure, component, or system which provide a means to control the quality of the material, structure, component, or system to predetermined requirements.

To augment this definition of quality assurance, here is Kendall and Buckland's (1971), p. 121, definition of quality control which is ...“the statistical analysis of process inspection data for the purpose of controlling the quality of a manufactured product which is produced in mass. It aims at tracing and eliminating systematic variations in quality, or reducing them to an acceptable level, leaving the remaining variation to chance. The process is said then to be statistically under control.”

Quality assurance is a two-way street—all products and services have both producers (shipper) and customers (receivers), although sometimes the producer and the customer are one and the same. Quality assurance criteria are constructed for specific purposes. Some criteria are designed to protect the consumer from a “raw deal” or from a health risk or from shoddy merchandise. Others are designed to protect the producer from excessive waste of time and material and overcrowded inventory. Still others are designed with both the consumer and the producer in mind, although sampling designs with such a balanced intention usually require great effort and resources and cooperation among the affected parties.

Do these criteria sound contradictory? They are not if both consumer and producer have the same quality assurance objectives in mind. Consider the

impacts of Japanese products on the American economy in the 1970s and 1980s. Quality was a major driving force: consumers wanted high-quality automobiles and electronics, and producers found ways to ensure that quality. Japanese industrialists understand quality assurance quite clearly. The enduring irony is that they learned it from W. Edwards Deming, a well-known American statistician. Deming tried to promote this philosophy to American industry during the 1940s and 1950s, but his message fell on the proverbial deaf ears.

Quality assurance procedures are data-based. The two main data types we are likely to encounter are continuous (such as weight or diameter) and discrete binary data (such as yes/no or in compliance/out-of-compliance). We have already met both types in the preceding chapters.

sampling by variables As we might expect, the practice of quality assurance has its own vocabulary. The treatment of continuous variables is called *sampling by variables*, while the treatment of discrete binary data is called *sampling by attributes*.

24.3 Process control: Building in quality

process control The claim that a product meets its promised specifications can be investigated during production, during post-production, or during both. The testing of product quality during production is called *process control*. Process control means that items are routinely sampled and checked to determine whether the production specifications are maintained during the product's manufacture. If the specifications are not met, then production is said to be out of control, and corrective actions are usually taken. These corrective actions can include a search for an assignable cause, an adjustment to the production process, and/or a retesting of the process before it is restarted. However, even though corrective actions may be taken to tighten (or adjust) the ongoing production process, the items produced before the adjustment may or may not be discarded.

Although process control is designed to guard against poor quality, it may also serve to indicate when the actual product is superior to that which is claimed by the manufacturer. This sometimes provides the manufacturer an opportunity to restate the product's specifications and to improve its competitive edge. At other times, the manufacturer may choose to relax some overly tight process controls, resulting in time and material savings, while still maintaining the claimed product specification.

24.4 Control charts for means

control charts Control of a production process is often monitored through *control charts*, which are graphic constructs that are designed to alert a quality control manager when a process is out of control—or, at least, appears to be heading in that direction. Duncan (1986), p. 417, summarizes the purposes and the consequences of using control charts:

A control chart is a statistical device principally used for the study and control of repetitive processes. Dr. Walter A. Shewhart, its originator, suggests that the control chart may serve, first, to define the goal or standard for a process that the management might strive to attain; second, it may be used as an instrument for attaining that goal; and, third, it may serve as means of judging whether the goal has been reached. It is thus an instrument to be used in specification, production, and inspection and, when so used, brings these three phases of industry into an interdependent whole.

Each control chart focuses on one target quantity—Shewhart’s “goal” or “standard”—such as the claimed mean, the advertised standard deviation or range, or the promised maximum proportion of defective items. A control chart for the mean, for instance, is constructed around the target quantity, which we designate by μ_{target} . The sample means of consecutively produced samples are plotted on the control chart with *control limits* constructed above and below the target value.

*control
limits*

Example 24.1 illustrates these ideas.

Example 24.1. Control chart for percentage of uranium in UO_2 powder. Jaech (1973), p. 68, gives an example involving the construction of a control chart for the average percent uranium in 19 batches of UO_2 powder. Table 24.1 reproduces this chart. Jaech assumes that the standard factor (i.e., the target mean, μ_{target}) is 87.60% and uses $\sigma = 0.06\%$, where σ is the standard deviation associated with each batch.

Table 24.1. Percent uranium in batches of UO₂ powder

Batch	1	2	3	4	5	6	7
Mean	87.54	87.56	87.50	87.47	87.64	87.56	87.71
Batch	8	9	10	11	12	13	14
Mean	87.61	87.60	87.60	87.47	87.60	87.69	87.78
Batch	15	16	17	18	19		
Mean	87.69	87.72	87.77	87.79	87.78		

upper control limit (UCL)

lower control limit (LCL)

Figure 24.1 shows the associated control chart. The horizontal centerline is plotted at the target value, $\mu_{target} = 87.60\%$. The **upper control limit (UCL)** is plotted at $\mu_{target} + 3\sigma = 87.60 + 3(0.06) = 87.78$, and the **lower control limit (LCL)** at $\mu_{target} - 3\sigma = 87.60 - 3(0.06) = 87.42$. The individual values of the 19 batches of UO₂ powder are plotted from left to right, in the sequence in which they are produced. The importance of ordering, usually by time, in control charts cannot be overemphasized. As we will see, evidence of an out-of-control process can be revealed by the ordering.

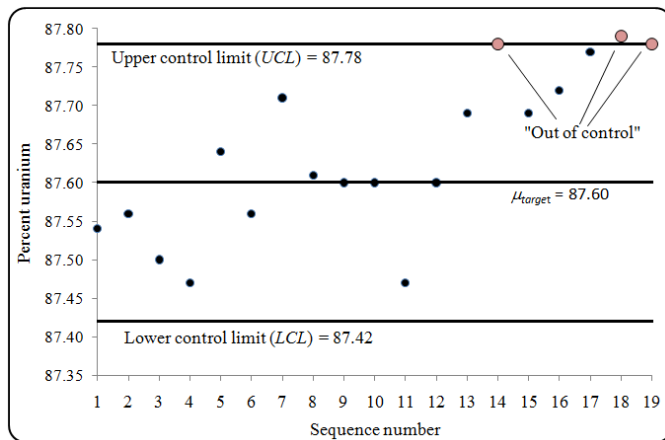


Figure 24.1. A control chart for batches of UO₂ powder (Table 24.1)

Note that, if the data values in Table 24.1 were based on multiple readings (i.e., on sample size $n > 1$), the control limits would be written as $87.60 \pm 3\sigma/\sqrt{n}$.

<i>Shewhart charts</i>	The selection of the value 3 as the coefficient of σ is consistent with long-standing practice in the field of quality control. Indeed, control charts were introduced as “3-sigma control charts” in Shewhart’s (1931) pioneering work on quality control; consequently, we may encounter references to 3-sigma control charts as <i>Shewhart charts</i> . Sometimes, we find the “3-sigma limits” referred to as <i>action limits</i> , a term designed to convey the urgency implied when observations fall beyond these limits.
<i>action limits</i>	
<i>warning limits</i>	As control chart practice developed in the United States, especially during and immediately after World War II, a seemingly endless procession of modified control limits appeared. For instance, <i>warning limits</i> (sometimes called <i>alarm limits</i> or <i>2σ limit</i>) may be set at $\mu_{target} \pm 2\sigma/\sqrt{n}$, with the implication of a warning—but with somewhat less urgency than action limits. The generic terminology refers to “ <i>k</i> -sigma limits” which are set at $\mu_{target} \pm k\sigma/\sqrt{n}$.
<i>alarm limits</i>	
<i>2σ limit</i>	

24.5 Run test for control charts

When we see a point lying outside the control limits—that is, the 3-sigma limits—the time has come to act. But how should we react to a point lying outside the 2-sigma limits, the “warning limits”? Do we just note the warning? Do we call in the manager? Do we have a committee meeting? Do we make a public announcement? Do we inform the regulating agency? The answer, depending on the production process, may be all, some, or none of the above.

The most important thing to do is to watch what happens on the chart as successive points are plotted. Do they fall between the warning limits, or do they continue to fall outside? If the latter, how do we react to points falling between the warning and the action limits?

Particularly powerful guidance, validated by decades of practice, can be found in a special area of statistics, “run theory.” Run theory finds application in many quantitative problems. Early references include Stevens (1939), Wald and Wolfowitz (1940), and Swed and Eisenhart (1943). Attention here, however, is paid to its application to control charts. Brownlee (1965), p. 224, and Duncan (1986), p. 417, make the necessary connections and guide the following discussion.

run Duncan (1986), p. 428, defines a **run** as a succession of items of the same class. Thus, suppose we have a collection of nine items: three of class *A*, two of class *B*, and four of class *C*. If we draw them at random, one at a time, without replacement, we might get the sequence {*BAAACCABC*}, thus providing two runs of *A* (one of length 2 and one of length 1), two runs of *B* (each of length 1), and two runs of *C* (one of length 3 and one of length 1).

With respect to control charts, we may look for several types of runs, such as outside certain *k*-sigma limits or above and/or below the target line. Indeed, in some situations, we might be on guard against “too many runs,” a situation that can develop when an inspector is helping us make “things average out.”

Here is a paraphrase of Duncan’s criteria (1986), p. 434, for suggesting an out-of-control condition in our process:

1. One or more points outside the control limits.
2. One or more points in the vicinity of a warning limit. A recommended step is the immediate collection and analysis of additional data to determine the actual level of production.
3. A run of seven or more points. Such a run may simply be above or below the target line. Or it may be a “run up” (a succession of increases in value) or a “run down” (a succession of decreases in value).
4. Cycles or other nonrandom patterns. Because of their infinite variety, such cycles or patterns are difficult to specify. Watch the charts—we’ll know them when we see them.
5. A run of two or three points outside the 2-sigma limits.
6. A run of four or five points outside the 1-sigma limits.

To make these matters explicit, consider Figure 24.2, which is a re-rendering of the control chart in Figure 24.1.

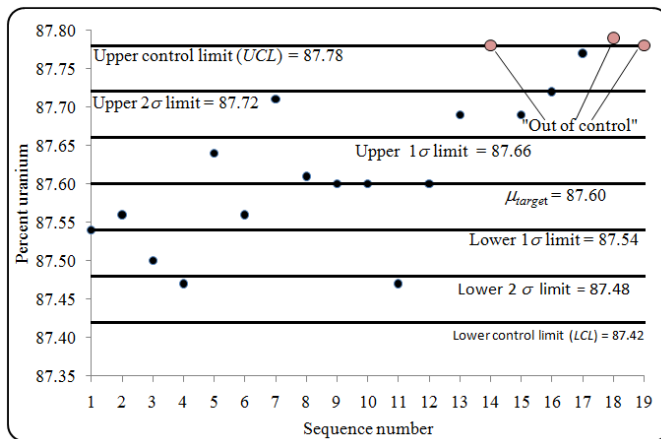


Figure 24.2. Figure 24.1 with σ and 2σ limits added

Starting with Batch 1, we have a run up of length 1, immediately followed by a run down of length 2. Only short runs occur until Batch 11, after which we see a run up of length 3. Indeed, Batch 14 is right on the UCL. Moreover, Batches 13 through 16 or 17 provide the alerting four or five points outside the 1-sigma limits. So we can see that, irrespective of Batch 14 being on the UCL, beginning with Batch 13, an out-of-control story is being told.

This type of application of run theory applies to all types of control charts, not just the means charts, one of which is used here for illustration.

- Some control chart practitioners, particularly those in Great Britain, use multipliers of 3.09 for control limits and of 1.96 for warning limits, instead of the 3σ and 2σ limits in Figures 24.1 and 24.2. These multipliers are the exact numbers associated with the 0.999 and 0.975 quantiles of the normal distribution.

24.6 Control charts for variability

In contrast to Example 24.1, most points plotted on a control chart are statistics derived from a sample of size $n > 1$. Each sample can therefore yield a measure (indeed, measures) of its own dispersion. Among these measures are the usual suspects: the sample variance, the sample standard deviation, and the sample range. Quality assurance practitioners use each of these according to the particular needs of the process being monitored.

A claim of “good production quality” requires not only a good average performance but also good consistency, or small variability, among the items being produced. To this end, some measure of within-sample dispersion

variability of the samples is subjected to control charts and examined for its being “in control” or “out of control” in the same way that means are evaluated.

Suppose we decide to apply control chart procedures to the standard deviations of the samples we’re using for controlling the mean. Suppose further that we have chosen a target measure of variability, denoted σ_{target} . We then construct a control chart for the sample standard deviations that is “anchored” on the target standard deviation, σ_{target} .

Our next task is to specify upper and lower control limits. Because the standard deviation is the square root of the variance, these control limits are based on the fact that the statistic $(n - 1)S^2/\sigma^2$ is distributed as a chi-square variable with $v = n - 1$ degrees of freedom, where S^2 is the sample variance and σ^2 is the population variance. To construct upper and lower control limits that correspond to the k -sigma limits associated with a tail probability α , we do the following:

- Step 1.** Obtain the value $\chi^2_{(1 - \alpha/2)}(n - 1)$ from Table T-2 of the appendix. Alternatively, obtain this value from Excel’s function =CHIINV($\alpha/2$, v).
- Step 2.** Equate the quantile from Step 1 to $(n - 1)S^2/\sigma^2_{target}$ and solve for S^2 in terms of σ^2_{target} . This yields the upper control limit, σ^2_{UCL} , for the variance.
- Step 3.** Obtain the value $\chi^2_{(\alpha/2)}(n - 1)$ from Table T-2. Alternatively, obtain this value from Excel’s function =CHIINV($1 - \alpha/2$, v).
- Step 4.** Equate the quantile from Step 3 to $(n - 1)S^2/\sigma^2_{target}$ and solve for S^2 in terms of σ^2_{target} . This yields the lower control limit, σ^2_{LCL} , for the variance.
- Step 5.** The control limits for the standard deviation are σ_{UCL} and σ_{LCL} .

The five steps above are written for a two-sided control limit. If we require a one-sided control limit, then:

- Use Steps 1, 2, and 5, with $\chi^2_{(1 - \alpha)}(n - 1)$, to obtain σ^2_{UCL} .
- Use Steps 3, 4, and 5, with $\chi^2_{(\alpha)}(n - 1)$, to obtain σ^2_{LCL} .

The upper and lower warning limits σ_{UWL} and σ_{LWL} for σ are constructed similarly to the upper and lower control limits with an appropriate value of α .

We illustrate these five steps in Example 24.2.

Example 24.2. Upper control and warning limits limits for σ .

Suppose we wish to verify the assumption in Example 24.1 that the standard deviation of the process is 0.06%. Construct a 99.5% upper control limit σ_{UCL} and a 95% upper warning limit σ_{UWL} , assuming that daily samples of size $n = 10$ are available for the calculation of S .

Step 1. From Table T-2, we obtain $\chi^2_{(0.995)(9)} = 23.6$ and $\chi^2_{(0.95)(9)} = 16.9$.

Step 2. We equate each of the quantiles so obtained to $(n - 1)S^2/\sigma^2_{target} = 9S^2/(0.06)^2$ and solve for S^2 .

We set $9S^2/(0.06)^2 = 23.6$ and solve for S^2 to obtain $S^2 = 0.0094$.
Hence, $\sigma^2_{UCL} = 0.00944$.

We set $9S^2/(0.06)^2 = 16.9$ and solve for S^2 to obtain $S^2 = 0.0069$.
Hence, $\sigma^2_{UWL} = 0.00676$.

Step 3. (not required)

Step 4. (not required)

Step 5. $\sigma_{UCL} = \sqrt{0.00944} = 0.0972$ and $\sigma_{UWL} = \sqrt{0.0069} = 0.0822$.

Figure 24.3 shows the control chart for σ^2 . Note that the sequence number is arbitrary.

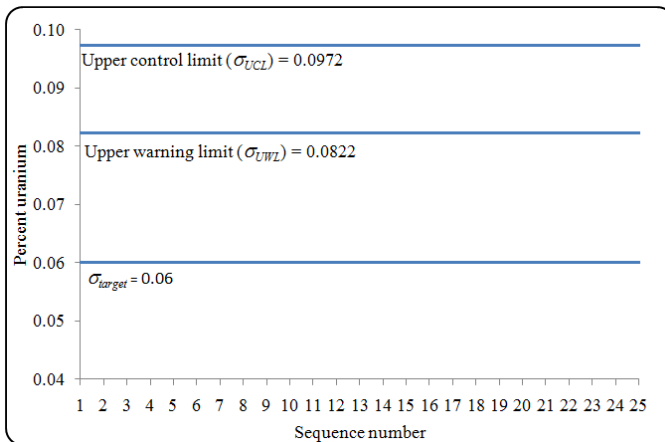


Figure 24.3. Control chart for standard deviation

24.7 Control charts for attributes

Consider next a process control where the specification of proportion of defective items in the lot or population is set at $\pi = \pi_{target}$. If the sample size is large enough that both $n(\pi_{target})$ and $n(1 - \pi_{target})$ are larger than 5, then a normal approximation to the distribution of the sample proportion is reasonable, as discussed in Chapter 22. The control chart is then constructed with a centerline at π_{target} , and the LCL and UCL may be set at:

$$\pi_{LCL} = \pi_{target} - 3\sqrt{\frac{\pi_{target}(1 - \pi_{target})}{n}}, \quad \pi_{UCL} = \pi_{target} + 3\sqrt{\frac{\pi_{target}(1 - \pi_{target})}{n}} \quad (24.1)$$

Lower and upper warning limits may be similarly constructed:

$$\pi_{LWL} = \pi_{target} - 2\sqrt{\frac{\pi_{target}(1 - \pi_{target})}{n}}, \quad \pi_{UWL} = \pi_{target} + 2\sqrt{\frac{\pi_{target}(1 - \pi_{target})}{n}} \quad (24.2)$$

As an example, if the target value of $\pi = \pi_{target} = 0.278$ and $n = 100$, then control limits for π are:

$$0.278 \pm 3\sqrt{\frac{(0.278)(0.722)}{100}} = 0.278 \pm 0.134$$

from which $\pi_{LCL} = 0.278 - 0.134 = 0.144$ and $\pi_{UCL} = 0.278 + 0.134 = 0.412$.

Further reading about process control in general, and control charts in particular, may be found in Burr (1976, 1979) and Duncan (1986).

24.8 Acceptance sampling: Verifying quality

Statistical quality assurance activities conducted after items are out of production (indeed, often after delivery is made to the consumer) fall into the domain of **acceptance sampling**. The basic idea of acceptance sampling is simple: A collection of items, usually called a **lot**, is examined, and the decision then is made to either accept or reject the entire collection—the lot—on the basis of that examination. The consequences of a rejection—scrapping the product, reworking the product to improve it, bankruptcy of the manufacturer—are too multifarious to explore here. Suffice it to say, rejection of any manufactured product is seldom taken lightly. It is the practice

**acceptance
sampling
lot**

of acceptance sampling that's difficult; convincing a group of managers that a lot of material must be rejected often requires the sharpest of diplomatic skills. Acceptance sampling follows an explicit and strict protocol that culminates in the acceptance or the rejection of an entire lot of items.

- In most acceptance-sampling literature, a "lot" is defined in terms similar to those used by Kendall and Buckland (1971), p. 88, as "... a group of units of a product produced under similar conditions and therefore, in a sense, of homogeneous origin; e.g., a set of screws produced by a lathe or a set of light bulbs produced by a number of similar machines. It is sometimes implicit that a lot is for inspection." Thus, "lot" fits the meaning of "population" as used in this book.

The value of acceptance sampling to the entire production process is captured in these excerpts from Duncan (1986), pp. 161–162:

It is to be emphasized that the purpose of acceptance sampling is to determine a course of action, not to estimate lot quality....

It is also to be emphasized that acceptance sampling is not an attempt to "control" quality. The latter is the purpose of control charts....

The indirect effects of acceptance sampling on quality are likely to be much more important than the direct effects.... Acceptance sampling...indirectly improves quality...through its encouragement of good quality by a high rate of acceptance and its discouragement of poor quality by a high rate of rejection.

Furthermore, if acceptance sampling is used ... at various stages of production, it may have beneficial effects in general on the quality of production.... Production personnel will ... become quality conscious and there will be an interest in quality on the part of both inspection and production [personnel]. The rule will be: Make it right the first time.

Duncan concludes his discourse by indicating the conditions under which acceptance sampling is likely to be used:

1. When the cost of inspection is high and the loss from the passing of a defective item is not great. It is possible in some cases that no inspection at all will be the cheapest plan.

2. When 100 percent inspection is fatiguing and a carefully worked-out sampling plan will produce as good or better results. [Because]...100 percent inspection may not mean 100 percent perfect quality,...the percentage of defective items passed may be higher than under a scientifically designed sampling plan.
3. When inspection is destructive. In this case sampling must be employed.

consumer's risk

Acceptance sampling is essentially a test of a hypothesis in which a sample statistic is compared against a claimed product's characteristic (such as a mean or a standard deviation). The key feature of acceptance sampling is to protect the consumer by limiting the *consumer's risk* through controlling the probability of accepting poor quality material. This translates into a small probability of a Type II error. In this context, a Type II error is made when we accept a "bad" lot. For further reading on acceptance sampling, refer to Duncan (1974), Burr (1976), or Schilling (1982).

Just as process control differentiates between sampling by attributes and sampling by variables, so does acceptance sampling. Here, we focus on sampling by attributes and assume that the attribute is undesirable, such as an item's being defective, broken, scratched, or not in compliance. Suppose further that the product's specification claims the proportion of items in the lot with the attribute is not larger than a quantity we denote by π_{target} .

To be absolutely certain sure that the fraction of the lot with the attribute of concern does not exceed π_{target} , we may need to examine every item in the lot. Such an effort is usually an unacceptable burden. Furthermore, for some products, such as sealed systems or single-use items, such inspection is impossible because the item's integrity is destroyed by the very act of inspection.

Statistical sampling allows us to inspect only a fraction of each submitted lot, while providing some assurance—in lieu of an absolute guarantee—that the proportion of items with the attribute meets the stated specification.

Quality assurance statements about a lot sampled for attributes convey a sense similar to these:

- *We are 95% assured that at least 95% of the items in the lot are in compliance.*
- *We are 95% confident that at least 95% of the cables in a bundle are traceable to their source.*

- *We are 99% confident that at least 99% of the pipe supports are properly welded.*
- *We are 90% certain that at least 80% of the invoices were paid within 30 days of their receipt.*

But beware! The last statement is not equivalent to:

- *We are 80% certain that at least 90% of the invoices were paid within 30 days of their receipt.*

Suppose we are given a lot of reinforcing bars (rebars), which was subjected to statistical sampling for meeting strength specifications. The quality statement issued for our review states: “We are 95% confident that at least 90% of the rebars in the lot are acceptable.” This specific criterion is often given the shorthand notation of 95/90. Thus, the first number (95) designates the assurance, and the second number (90) designates the quality.

assurance-to-quality As a mnemonic for keeping track of these two values, consider this as being an *assurance-to-quality* statement and denote it by the symbol A/Q . In many inspection programs in general, and in NRC programs in particular, the A/Q statement is set at 95/95. Thus, we focus here on the 95/95 criterion to illustrate the technique for sampling.

24.9 The A/Q criterion: Rules of the game

The general A/Q criterion is accompanied by a strict set of rules. Stick to these rules and avoid distracting debate. Changes to the rules should be considered only on the conservative side and because of compelling arguments. These rules are:

- The sample size must be determined before sampling begins and the entire lot will be either accepted or rejected, depending on the results of the collected sample.
- The lot must be unequivocally defined before sampling begins. It is inappropriate to redefine the lot once the inspection begins. Note, however, that the lot size does not necessarily have to be known.
- The lot is made up of similar items that are treated alike (i.e., they are interchangeable in terms of their intended use).
- If the lot is comprised of several sub-lots, we must address each sub-lot separately.

- The A/Q criterion is expressed as two numbers (e.g., 95/95, 90/95, 95/99). The first number (A) is the assurance that the acceptable proportion is met. The second number (Q) is the acceptable percentage of “good” items.
- The assurance level (A) is a lower bound on the probability that the lot will not be accepted when the lot's quality is less than Q , e.g., when the percentage of defective items in the lot is greater than $100(1 - Q)\%$. As such, A provides a “comfort level” for the consumer.
- The A/Q criterion provides the designed assurance to the consumer, but it is not designed to provide assurance to the producer. The producer, however, may wish to know the probability that the lot will be accepted when the quality of the lot is at least as good as the product specification, Q . This information is provided by the operating characteristic (OC) of the plan under consideration. This OC is similar to the operating characteristic curve for testing a hypothesis. See, for example, Duncan (1986), p. 363, or Schilling (1982).
- Once we select a sampling plan and begin sampling, we do not switch plans. We will see, by way of example at the end of this chapter, how changing horses in midstream is likely to cause a deterioration of the promised assurance.

The calculation of probabilities associated with the A/Q process is based on the binomial distribution, the requirements of which are detailed in Chapter 22.

For any specified values of A and Q , the sample size n required to satisfy the A/Q criterion is based on the probability of drawing a sample with a specified number of defective items from the lot. Because a lot contains a finite number of individual items, this probability is governed by the hypergeometric distribution (see Chapter 21). However, if the sample size is considerably smaller than the lot size N (say, $n < N/10$), then the binomial approximation to the hypergeometric distribution may be used (see Section 22.11). Moreover, because the binomial approximation requires a larger sample size than does the hypergeometric distribution, the actual assurance level will be higher than the design level. If the sample size is not small compared to the lot size, then the hypergeometric distribution should be used for the probability calculations. In this case, see Sherr (1973) for the required sampling plan.

acceptance number

Given A and Q , the sampling plan determines the sample size, denoted by n , and the maximum number of items with the attributes that can be tolerated in the sample. This maximum is called the **acceptance number**, and is denoted here by c . If the sample yields more than c defective items, the entire lot is rejected.

Because the sample size is fixed in advance, this plan is called a single-sampling plan. Once the number of defectives in the sample exceeds c , the lot must be rejected and there is no "second chance." However, double- or multiple-sampling plans may be designed (at the expense of a larger initial sample) to allow a second chance for acceptance if the number of defectives in the initial sample is not too large. For details about these special plans, see Duncan (1986, p. 184) or Schilling (1982, p. 127). The A/Q is illustrated by an example, presented in the next section.

24.10 Probability calculations for the 95/95 criterion

The focus here is on the $A/Q = 95/95$ criterion for quality assurance. Used routinely by the NRC, this criterion has a special role throughout the nuclear power industry.

There are many sampling plans that satisfy the 95/95 criterion. Three such plans are displayed side-by-side in Table 24.2, and are labeled as Plan A, Plan B, and Plan C.

Table 24.2. Three 95/95 sampling plans

Plan A	Plan B	Plan C
Select $n_A = 59$ items at random from the lot.	Select $n_B = 93$ items at random from the lot.	Select $n_C = 124$ items at random from the lot.
If 0 items fail, accept the lot.	If 0 or 1 item fails, accept the lot.	If 0, 1, or 2 items fail, accept the lot.
If even 1 unacceptable item is found, reject the lot.	If 2 or more unacceptable items are found, reject the lot.	If 3 or more unacceptable items are found, reject the lot.
For this plan, the acceptance number is $c = 0$.	For this plan, the acceptance number is $c = 1$.	For this plan, the acceptance number is $c = 2$.

To show that each of the three candidate plans meets the desired $A/Q = 95/95$ criterion, we need to calculate some binomial probabilities for $\pi = 0.05$. Because appendix Table T-8 is restricted to probabilities for $n \leq 40$, we constructed Table 24.3 to give the selected probabilities needed for our immediate task. These probabilities can be derived with the methodology developed in Chapter 22 or using Excel's =BINOMDIST($y, n, 0.05, 0$) for marginal probabilities and =BINOMDIST($y, n, 0.05, 1$) for cumulative probabilities.

Table 24.3. Selected binomial probabilities, $\pi = 0.05$

n	Marginal probabilities			Cumulative probabilities		
	$y = 0$	$y = 1$	$y = 2$	$y \leq 0$	$y \leq 1$	$y \leq 2$
58	0.0510	0.1558	0.2337	0.0510	0.2069	0.4406
59	0.0485	0.1506	0.2298	0.0485	0.1991	0.4289
⋮	⋮	⋮	⋮	⋮	⋮	⋮
92	0.0089	0.0432	0.1035	0.0089	0.0521	0.1556
93	0.0085	0.0415	0.1005	0.0085	0.0500	0.1504
⋮	⋮	⋮	⋮	⋮	⋮	⋮
123	0.0018	0.0118	0.0378	0.0018	0.0136	0.0514
124	0.0017	0.0113	0.0365	0.0017	0.0130	0.0495

To show that each of the three plans meets the 95/95 criterion, we convert the $A = 95\%$ assurance requirement to the complementary statement that the probability of accepting the lot is no larger than 0.05 when the proportion of good items in the lot is $Q = 95\%$. Because the acceptance number c is the number of defective items in the sample, the probabilities in Table 24.3 are the maximum cumulative probabilities of y defective items in the samples.

Using the same argument as for Plan A, we find that Plan C, for which $n = 124$ and $c = 2$, satisfies the 95/95 criterion.

We see that each of Plans A, B, and C satisfies the 95/95 criterion. If we must use one of these plans, which one should we choose and why?

From the consumer's perspective, it doesn't really matter since all three plans provide the same assurance. However, from the producer's perspective, there are two other considerations. The first is the cost of the sample and the second is the probability of a good lot being rejected. To minimize the sample size, the producer should opt for Plan A. However, if Plan A were chosen, then the lot would be rejected if even a single item out of 59 (that's less than 2% of the

sample!) is defective. If the producer is very confident that the lot quality is very high, say, 99%, he may choose to stay with Plan A. On the other hand, to protect himself if the lot quality is closer to 95%, he may prefer Plan C. With its larger sample size, Plan C is less likely to reject the lot with 3 defectives out of 124 sampled (a 2.4% rate) than Plan B (2 out of 93 or 2.2%) and certainly less likely than Plan A (1 out of 59 or 1.7%).

24.11 What's wrong with this picture?

As emphasized earlier, once we select a plan, we must stick with it. This section discusses the common mistake of changing plans after sampling begins. As an illustration, consider a hypothetical producer who agrees at the outset that each of the three sampling plans described in Table 24.2 satisfies the required 95/95 criterion. However, he sees nothing wrong with the following multiple-sampling strategy that he believes will still satisfy the 95/95 assurance criterion. This plan would lead to the acceptance of the lot if any of the following four events occurs:

- Event 1.** Use the approved Plan A. Collect 59 items. If zero defective items are found, stop the sampling and accept the lot.

- Event 2.** Use the approved Plan A. Collect 59 items. If exactly 1 defective item is found, collect 34 additional items. This brings the total number of inspected items to 93, as required for the approved Plan B. If no defective items are found among the additional 34 items, stop the sampling and accept the lot.

- Event 3.** Use the approved Plan A. Collect 59 items. If exactly 2 defective items are found, collect 65 additional items. This brings the total number of inspected items to 124, as required by the approved Plan C. If no defective items are found among the additional 65 additional items, stop the sampling and accept the lot.

- Event 4:** Use the approved Plan A. Collect 59 items. If exactly 1 defective item is found, collect 34 additional items, bringing the total number of items inspected to 93, as required for the approved Plan B. If exactly 1 defective item is found among the 34 items, collect 31 additional items, bringing the total number of items inspected to 124, as required for the approved Plan C. If no additional defective items are found, stop the sampling and accept the lot.

At first glance, the producer's multiple-sampling strategy appears reasonable and acceptable since each event satisfies one of the approved plans. But it always pays to revisit first principles and calculate the assurance associated with this sampling plan. To calculate that assurance, we start with some intermediate probability calculations. Although we could obtain at least some of those

probabilities from a table, it is instructive to calculate them and present all the necessary calculations together. Using Chapter 22 methodology, recall that:

$$Pr\{y, n | \pi\} = \frac{n!}{y! (n-y)!} \pi^y (1-\pi)^{n-y} \quad (24.3)$$

More specifically,

$$Pr\{y, n | \pi = 0.05\} = \frac{n!}{y! (n-y)!} 0.05^y 0.95^{n-y} \quad (24.4)$$

where $Pr\{y, n | \pi = 0.05\}$ denotes the probability that Y defective items will be found in a sample of size n , given that the proportion of defective items in the lot is $\pi = 0.05$. Relevant probabilities (some of which are already shown in Table 24.3) are given below:

$$Pr\{0, 31 | 0.05\} = 0.2039$$

$$Pr\{0, 34 | 0.05\} = 0.1748$$

$$Pr\{0, 59 | 0.05\} = 0.0485$$

$$Pr\{0, 65 | 0.05\} = 0.0356$$

$$Pr\{1, 31 | 0.05\} = 0.3327$$

$$Pr\{1, 34 | 0.05\} = 0.3128$$

$$Pr\{1, 59 | 0.05\} = 0.1506$$

$$Pr\{2, 59 | 0.05\} = 0.2298$$

Next, calculate the probability of occurrence of each of the four events listed earlier:

$$\begin{aligned} Pr\{\text{Event 1}\} &= Pr\{y = 0, n = 59\} \\ &= Pr\{0, 59\} = 0.0485 \end{aligned}$$

$$\begin{aligned} Pr\{\text{Event 2}\} &= Pr\{(y = 1, n = 59 \text{ and } y = 0, n = 34)\} \\ &= Pr\{1, 59\} Pr\{0, 34\} \\ &= (0.1506)(0.1748) \\ &= 0.0263 \end{aligned}$$

$$\begin{aligned}
 Pr\{\text{Event 3}\} &= Pr\{y = 2, n = 59 \text{ and } y = 0, n = 65\} \\
 &= Pr\{2, 59\} Pr\{0, 65\} \\
 &= (0.2298)(0.0356) \\
 &= 0.0082
 \end{aligned}$$

$$\begin{aligned}
 Pr\{\text{Event 4}\} &= Pr\{y = 1, n = 59 \text{ and } y = 1, n = 34, \text{ and } y = 0, n = 31\} \\
 &= Pr\{1, 59\} Pr\{1, 34\} Pr\{0, 31\} \\
 &= (0.1506)(0.3128)(0.2039) \\
 &= 0.0096
 \end{aligned}$$

Finally, put the pieces together. These four events are mutually exclusive; therefore, the probability of accepting the lot is the sum of the four probabilities. This is calculated as:

$$\begin{aligned}
 Pr\{\text{lot accepted}\} &= 0.0485 + 0.0263 + 0.0082 + 0.0096 \\
 &= 0.0926
 \end{aligned}$$

Hence, with 5% of the items being defective, the probability of accepting the lot is 0.0926—almost twice as large as the 0.05 that was intended. The probability of rejecting the lot is $1.0000 - 0.0926 = 0.9074$. Thus, the assurance is only 90.7%, not the claimed 95%.

The reason for the decrease in the assurance level is that the combination of the three 95/95 plans provides many more possibilities for lot acceptance. Because each of the three plans was designed to achieve the 95/95 criterion as closely as possible, any increase in the possibilities for lot acceptance without a change in the sample sizes or acceptance numbers necessarily leads to a decrease in the assurance level.

Of course, legitimate multiple-sampling plans can be constructed so that the plan assurance will be 95/95—but such plans are by no means unique. A discussion of multiple-sampling plans appears in, for example, Duncan (1986), pp. 204-213.

Finally, keep in mind that this chapter's acceptance sampling plans are based on infinite or at least "large" lots where the binomial distribution is used as the theoretical base. Small lots, however, are better served by sampling plans that are based on the hypergeometric distribution (Chapter 21). The design of sampling plans for lots governed by the hypergeometric distribution is far from trivial. A detailed treatment of A/Q procedures may be found in Duncan (1986), Burr (1979), or Schilling (1982).

25

Nonparametric statistics

25.1 What to look for in Chapter 25

Chapter 25 introduces statistical methods that examine data from populations with unknown or unspecified distributions. This chapter provides the following statistical tests that may be used when the distribution of the population is not known or cannot be assumed:

- *Runs test for randomness, §25.3*
- *Sign test for testing location for a single population, §25.4*
- *Wilcoxon signed ranks test for a single population, §25.5*
- *Sign test for testing locations with paired samples, §25.6*
- *Wilcoxon test for testing locations with paired samples, §25.7*
- *Wilcoxon test for comparing two locations, §25.8*
- *Median test for locations of several populations, §25.9*
- *Kruskal-Wallis test for locations of several populations, §25.10*
- *Rank ANOVA test for locations of several populations, §25.11*
- *Friedman test for locations of several populations, §25.12*
- *Squared ranks test for comparing two variances, §25.13*
- *k-sample squared ranks test for comparing variances, §25.14*
- *Spearman test of independence for two populations, §25.15*

25.2 Nonparametric methods

Many statistical tests presented in previous chapters, especially those involving continuous distributions, assume that the given samples come from a normal distribution or from another distribution with defined parameters. These distributions are called *parametric distributions* because the form of the distribution is assumed to be known. For such distributions, parameter estimation and tests of hypotheses are the domain of *parametric statistics*. However, if an assumption about the form of the distribution cannot be made, and if the central limit theorem is not applicable, we turn to *nonparametric statistics*. Nonparametric statistics, also called *distribution-free statistics*, is a sub-discipline of statistics that deals with procedures that do not require knowledge of the distribution that gave rise to the sample we wish to analyze.

Nonparametric methods have several advantages over parametric methods:

- There is no need to make assumptions about unknown populations. Thus, nonparametric methods offer protection against misleading conclusions that may be drawn because of incorrect distributional assumptions.
- Calculations are usually easier than calculations required by counterpart parametric methods.
- Because interval data are transformed to ordinal data (see Section 1.5), outlying observations are not given undue weight.

Nonparametric statistics methods, however, have one disadvantage:

- When the form of the population distribution is known, more information can be extracted from the sample. Specifically, nonparametric methods result in less sensitive hypothesis tests than parametric methods that take advantage of the assumed knowledge of the distribution.

There are a great many nonparametric procedures in the literature, most of which address specific data irregularities. In this chapter we present some of the most common procedures. More detailed discussions of nonparametric procedures may be found in Conover (1980), Hollander and Wolfe (1973), and Siegel and Castellan (1988).

Some of the material presented in this chapter is taken from Chapter 9 of Bowen and Bennett (1984).

25.3 Test of randomness: The runs test

runs test Distribution-free procedures are not assumption-free. Just as in the parametric case, we assume that the collected sample is drawn at random. However, it is possible to test this assumption with a nonparametric test of randomness. This test, called the *runs test*, is based on the order in which the observations were collected.

Consider a sequence of *n* observations recorded in the order in which they are collected. Replace each observation that is larger than the sample median by a plus sign “+” or by a minus sign “-” otherwise.

➤ Note that an observation that is equal to the median is counted as a minus.

The runs test tests for randomness by checking if there is a pattern to the sequence of observations. To motivate it, we examine three sequences of pluses and minuses:

Sequence 1 appears to be a random sequence, i.e., it has no pattern.

+ - - + + + - + - - - + - - + - - + + - - - + - - + + (Sequence 1)

If we see clustering of signs, such as in the extreme case of Sequence 2, we are very likely to question the randomness of the sequence.

- - - - - - - - - - - - + + + + + + + + + + + + + + + + + (Sequence 2)

We would also be skeptical if we see Sequence 3.

- + - + - + - + - + - + - + - + - + - + - + - + - + - + - + (Sequence 3)

To quantify the difference in the patterns of Sequences 1,2, and 3, we introduce the concept of a run, which is defined as a succession of identical signs. In Sequence 1 we have a run of 1 plus, followed by a run of 2 minuses, followed by a run of 3 pluses, etc., for a total of 17 runs. In Sequence 2 we have just 2 runs, and in Sequence 3 we have 30 runs.

➤ There is a similar concept in Section 24.5 -- the “run test” (note the singular form), a procedure for detecting irregularity used in quality assurance.

Let *U* denote the number of runs in the ordered sample. With a predetermined level of significance *α*, we use the runs test to check whether *U* is excessively

large or excessively small for a sample of size n . The test of hypothesis is set up as:

$$H_0: \text{Sequence is random} \quad (25.1)$$

$$H_1: \text{Sequence is not random} \quad (25.2)$$

Let n_1 denote the smaller of the number of pluses and the number of minuses, or either number if they are equal. Let $n_2 = n - n_1$ be the number of minuses. For $n_1 \leq 10$ and $n_2 \leq 10$, Swed and Eisenhart (1943) calculated the cumulative probabilities $Pr\{U \leq u\}$ for obtaining a number of runs as high as u . A table of these probabilities, compiled by Draper and Smith (1981), is given in Table T-15 of the appendix.

The null hypothesis is rejected if:

$$Pr\{U \leq u\} \leq \alpha/2 \quad (25.3)$$

- We are not aware of any parametric test that tests the chronological randomness of a sequence. From Gibbons (1976), p. 365, "There is no parametric test that tests the chronological randomness of a sequence."

Example 25.1. Percent of uranium in UO_2 powder (Bowen and Bennett (1988), p. 486)). To estimate an average percent by weight uranium value for a production lot of UO_2 powder, we analyze 14 aliquots drawn as the powder is produced. The percent uranium values (in the order of production) are listed in Table 25.1. The median of the 14 values is 87.600, the average of the middle two observations, 87.599 and 87.601. The (+) and (-) signs below the observations are assigned according to the rule for the runs test. Thus, the first three values are less than 87.600 and are replaced with minus signs, while the fourth and fifth values are greater than 87.600 and are replaced with plus signs.

Table 25.1. Percent of uranium in UO_2 powder, in their order of collection

| | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|
| 87.590 | 87.588 | 87.595 | 87.603 | 87.601 | 87.582 | 87.580 |
| (-) | (-) | (-) | (+) | (+) | (-) | (-) |
| 87.597 | 87.599 | 87.609 | 87.601 | 87.610 | 87.611 | 87.615 |
| (-) | (-) | (+) | (+) | (+) | (+) | (+) |

This process yields the sequence:

- - - + + - - - + + + +

(Sequence 4)

We test the hypothesis that these observations are a random sample from the production lot using the runs test with $\alpha = 0.05$. In (Sequence 4) we have $n_1 = 7$ plus signs and $n_2 = 7$ minus signs, with $u = 4$ runs. From Table T-15, we have $Pr\{U \leq 4\} = 0.025$.

Because $Pr\{U \leq 4\} = 0.025 \leq 0.05/2 = 0.025$, the null hypothesis of randomness is rejected at the $\alpha = 0.05$ [level of significance]. Thus, fewer runs were observed than would be expected if H_0 were true, and the sample is not regarded as a random sample. The implication of lack of randomness is important from an accountability viewpoint, because an average percent uranium for the lot has less meaning when the observations are not a random sample from the production lot.

A table for testing H_0 when $n_1 \leq 20$ and $n_2 \leq 20$ can be found in Table G by Siegel and Castellan (1988), which is modified from Swed and Eisenhard (1943). That table, however, is restricted to $\alpha = 0.05$.

When $n_1 > 10$ and $n_2 > 10$, exact critical values for the runs test are not provided because a normal approximation to the distribution of U is satisfactory. In such cases, the statistic

$$Z = \frac{U - \mu + a}{\sigma} \quad (25.4)$$

is approximately distributed as a standard normal variable, where μ and σ are the mean and standard deviation of the discrete distribution of U under the null hypothesis of randomness. These parameters are (Siegel and Castellan (1988), p. 62):

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad (25.5)$$

$$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \quad (25.6)$$

The term a in Equation (25.4) is a continuity correction that helps to compensate for a continuous distribution used to approximate a discrete distribution. It is equal to:

$$a = \begin{cases} 0.5, & \text{when testing for too few runs} \\ -0.5, & \text{when testing for too many runs} \end{cases} \quad (25.7)$$

A two-sided test of randomness is performed by comparing the test statistics $Z_{100 \text{ many}}$ and $Z_{100 \text{ few}}$ against the standard normal quantiles $z_{\alpha/2}$ and $z_{1-\alpha/2}$ obtained

from Table T-1 of the appendix. The null hypothesis of randomness is rejected if either $Z_{\text{too many}} > z_{1-\alpha/2}$ or $Z_{\text{too few}} < z_{\alpha/2}$.

Example 25.2. Runs test for a large sample. For a problem similar to that in Example 25.1, analyses on 25 aliquots were performed with $n_1 = 12$, $n_2 = 13$, and $U = 17$. We test the hypothesis of randomness at the $\alpha = 0.05$ level of significance.

The normal approximation is used with the test statistic Z computed from Equation (25.4), where μ is obtained from Equation (25.5) and σ from Equation (25.6) as:

$$\mu = \frac{2(12)(13)}{12 + 13} + 1 = 13.48$$

$$\sigma = \sqrt{\frac{2(12)(13) \cdot 2(12)(13) - 12 - 13}{(12 + 13)^2 (12 + 13 - 1)}} = \sqrt{5.97} = 2.44$$

Hence:

$$z_{\text{too many}} = \frac{17 - 13.48 - 0.5}{2.44} = 1.24$$

$$z_{\text{too few}} = \frac{17 - 13.48 + 0.5}{2.44} = 1.65$$

The quantiles $z_{0.025} = -1.960$ and $z_{0.975} = 1.960$ are obtained from Table T-1. Because $z_{\text{too few}} = 1.65 > -1.960$, and $z_{\text{too many}} = 1.24 < 1.960$, the null hypothesis of randomness is not rejected. Thus, there is no statistical evidence for a non-random sequence of the data.

25.4 Test of location: The sign test

In Chapter 13 we tested the mean of a population from a normal distribution. When the variance is not known (the most common case) we use Student's t -statistic for the test (Section 13.7). Nonparametric procedures refer to tests of centrality as tests of location, with the center of the population typically being the median rather than the mean. If the population is symmetric, the median is equal to the mean.

This chapter presents two procedures for location testing -- the sign test in this section and the Wilcoxon signed rank test in the next section. Both procedures test whether the sample comes from a population with median ξ_0 , using α as the level of significance. In both procedures the null and alternative hypotheses are written as:

$$H_0: \xi = \xi_0 \quad (25.8)$$

$$H_1: \xi \neq \xi_0 \quad (25.9)$$

where ξ is the Greek letter xi (pronounced ksi).

The **sign test** assumes that a random sample is drawn from a single population and that the characteristic of interest is measured on at least an ordinal scale. Each observation is replaced with a “+” if the observation is larger than ξ_0 , and with a “-” if it is smaller than ξ_0 . Observations equal to ξ_0 are discarded and the sample size is reduced accordingly. Let n denote the reduced sample size.

The test statistic for the sign test is T which is the number of plus signs. We run a two-sided test to determine whether T is too large or too small compared to the number that is expected when H_0 is true.

Under H_0 , the test statistic T has a binomial distribution with $\pi = 0.50$, as 50% of the observations are expected to be above and 50% below the median ξ_0 . (See Section 22.3.)

Let t be the observed value of T . For significance level α , calculate the tail probability Q that T is equal to t or a more extreme value, and reject H_0 if Q is too small. Specifically, reject H_0 if:

$$Q \leq \alpha/2 \quad (25.10)$$

Using Equation (22.4), the formula for Q is a cumulative probability:

$$\text{If } t < n/2, \text{ then } Q = \sum_{i=0}^t \frac{n!}{i!(n-i)!} (0.5)^n \quad (25.11)$$

$$\begin{aligned} \text{If } t > n/2, \text{ then } Q &= \sum_{i=t}^n \frac{n!}{i!(n-i)!} (0.5)^n \quad (25.12) \\ &= \sum_{i=0}^{n-t} \frac{n!}{i!(n-i)!} (0.5)^n \end{aligned}$$

- The above equations do not include the case where $t = n/2$ because, with half the pluses above and half below ξ_0 , we do not reject H_0 .

Excel's functions =BINOMDIST($t, n, 0.5, 1$) and =BINOMDIST($n - t, n, 0.5, 1$) return the values of the summations in Equations (25.11) and (25.12), respectively.

For small samples, the required summations may be calculated by summing the marginal probabilities under $\pi = 0.50$ in Table T-8 of the appendix.

Example 25.3. Bias in solid waste measurement (Bowen and Bennett, 1988, p. 490). A non-destructive assay (gamma scan) instrument is checked for bias by measuring the ^{235}U (in grams) in a solid waste barrel working standard with an assigned value of 2.40 g. During a six-month period, the standard is measured 12 times with the results given in Table 25.2. Test the hypothesis that the population median is equal to 2.40 g. Use the sign test with $\alpha = 0.05$ level of significance.

Table 25.2. Grams of ^{235}U in a solid waste barrel working standard

| | | | | | |
|------|------|------|------|------|------|
| 2.92 | 2.35 | 2.44 | 2.21 | 2.58 | 2.65 |
| (+) | (-) | (+) | (-) | (+) | (+) |
| 2.73 | 2.04 | 2.42 | 2.62 | 2.17 | 2.39 |
| (+) | (-) | (+) | (+) | (-) | (-) |

The plus or minus signs in parentheses indicate whether the observation is above or below the hypothesized median of 2.40.

In this example we have $n = 12$ and $t = 7$. Since $t > n/2 = 6$, we use Equation (2.12) and Table T-8. With $\pi = 0.50$, $n - t = 5$, and $y = 0, 1, 2, 3, 4, 5$, we have $Q = 0.0002 + 0.0029 + 0.0161 + 0.0537 + 0.1208 + 0.1934 = 0.3871$, rounded to 0.387.

Alternatively, we can use Excel's binomial function, =BINOMDIST(5, 12, 0.5, 1) which returns the value of 0.3872, rounded to 0.387.

Because $0.387 > \alpha/2 = 0.025$ we do not reject H_0 . We have no statistical evidence to claim that the population median is different from 2.40.

In the parametric counterpart of Example 25.3, we would assume that the sample comes from a normal distribution with unknown standard deviation and test $H_0: \mu = 2.40$ vs. $H_1: \mu \neq 2.40$. From Section 13.7, we use a t -test based on the sample mean $\bar{y} = 2.460$ and sample standard deviation $s = 0.253$. Using Equation (13.9), Student's t is calculated as:

$$t = \frac{2.460 - 2.40}{0.253 / \sqrt{12}} = 0.822$$

From Table T-3 of the appendix, the critical value for a two-sided test with $\alpha = 0.05$ is $t_{0.975}(11) = 2.20$. Because $|0.822| < 2.20$, we have no statistical evidence against H_0 , thus reaching the same conclusion as the sign test.

When n is large, we could use the normal approximation to the binomial distribution for testing whether the median is equal to ξ_0 with the sign test. To see how large n needs to be, recall from Equation (22.10) that the normal approximation to the binomial is adequate if both $n\pi \geq 5$ and $n(1 - \pi) \geq 5$. Since $\pi = (1 - \pi) = 0.50$, this means that the normal approximation can be used if $n \geq 5/0.5 = 10$.

To use the normal approximation to the binomial for a sample of size n , we use the larger of the following two Z statistics (see Gibbons (1976), Chapter 3, Equation (2.2)):

$$Z = \frac{T - 0.5 - 0.5n}{0.5\sqrt{n}} \quad (25.13)$$

$$Z = \frac{(n - T) - 0.5 - 0.5n}{0.5\sqrt{n}} \quad (25.14)$$

where the term 0.5 is the correction for continuity.

Example 25.4. Normal approximation to the sign test. To verify the accuracy of a scale, a 1.077 kg standard was weighed $n = 30$ times on the scale. There were 8 observations below and 22 above 1.077 kg. Use the sign test to test whether the scale is biased, with $\alpha = 0.05$.

The null hypothesis is $H_0: \xi_0 = 1.077$. The larger of the two Z values in Equations (25.13) and (25.14) is

$$z = \frac{22 - 0.5 - 0.5(30)}{0.5\sqrt{30}} = 2.37$$

The critical value for this test is $z_{1-\alpha/2} = z_{0.975} = 1.960$. Since $2.37 > 1.960$, we conclude that the scale is biased.

25.5 Test of location: Wilcoxon signed ranks test

The Wilcoxon signed ranks test is another test of location for a single population. The test assumes that the distribution is symmetric about its median, ξ , which is consequently equal to its mean. The null hypothesis is $H_0: \xi = \xi_0$. The alternative test may be two-sided, ($H_1: \xi \neq \xi_0$) or one-sided ($H_1: \xi > \xi_0$ or $H_1: \xi < \xi_0$).

The Wilcoxon signed ranks test considers only observations that are not equal to ζ_0 . Let y_1, y_2, \dots, y_n be a (reduced) random sample of n observations of the random variable Y , where the characteristic of interest is measured on at least an ordinal measurement scale.

For each y_i , $i = 1, 2, \dots, n$, calculate $d_i = \zeta_0 - y_i$ and its absolute value $|d_i|$. Then order the $|d_i|$ in ascending order and rank them d from 1 to n , with the smallest $|d_i|$ having rank 1 and the largest rank n . If two or more of the $|d_i|$ are equal, assign each the average of the ranks that would otherwise have been assigned. For example if the eighth and ninth ordered $|d_i|$ are equal, each is given the rank of 8.5. Similarly, if the 12th, 13th, and 14th ordered $|d_i|$ are equal, each of the three is given the rank of 13.

- The ranking of a group of numbers may be done by calling on Excel's =RANK(number, reference, order). Place the cursor where you wish the rank to print. The first entry in parenthesis is the value you wish to rank, "reference" is the range of the values to be ranked, and "order" is any non-zero number if the ranking is done in increasing order of magnitude, and 0 otherwise. For example, =RANK (A7, A1..A10, 1) returns the rank of the value in cell A7 if the values were ordered in increasing order of magnitude.

Let $R_i = 0$ if d_i is negative and $R_i =$ the rank of $|d_i|$ if d_i is positive. (Note that no $d_i = 0$.) The test statistic for the Wilcoxon signed ranks test is:

$$W = \sum_{i=1}^n R_i \quad (25.15)$$

In other words, W is the sum of ranks for the positive d_i . A value of W that is either too small or too large leads to the rejection of the null hypothesis. Specifically, when $H_0: \zeta = \zeta_0$:

$$\text{if } H_1 \text{ is } \zeta \neq \zeta_0, \text{ reject } H_0 \text{ if } W > w_{1-\alpha/2} \text{ or if } W < w_{\alpha/2} \quad (25.16)$$

$$\text{if } H_1 \text{ is } \zeta > \zeta_0, \text{ reject } H_0 \text{ if } W > w_{1-\alpha} \quad (25.17)$$

$$\text{if } H_1 \text{ is } \zeta < \zeta_0, \text{ reject } H_0 \text{ if } W < w_{\alpha} \quad (25.18)$$

where w_q is obtained from Table T-16 of the appendix for $n \leq 50$ and $q \leq 0.50$. For $q > 0.50$ use $w_q = n(n+1)/2 - w_{1-q}$. Table T-16 provides the values $n(n+1)/2$ in the column on the right.

For $n > 50$, quantiles for w_α may be approximated by:

$$w_\alpha = n(n+1)/4 + z_\alpha \sqrt{n(n+1)(2n+1)/24} \tag{25.19}$$

where $z_\alpha = 1 - z_q$ and z_q is the q^{th} quantile of the standard normal distribution, obtainable from Table T-1 of the appendix.

Example 25.5. Bias in solid waste measurement (continued). Use the Wilcoxon signed ranks test on the data from Table 25.2 (Example 25.3) to test whether $\xi = \xi_0$, with $\alpha = 0.05$ as the level of significance.

Table 25.3 lists the ordered observations y_i , the observed differences $d_i = 2.40 - y_i$, the ranks of the $|d_i|$, and the values r_i .

Table 25.3. Wilcoxon signed ranks test

| y_i | $d_i = 2.40 - y_i$ | Rank of $ d_i $ | r_i |
|-------|--------------------|-----------------|-------|
| 2.04 | 0.36 | 11 | 11 |
| 2.17 | 0.23 | 8 | 8 |
| 2.21 | 0.19 | 6 | 6 |
| 2.35 | 0.05 | 4 | 4 |
| 2.39 | 0.01 | 1 | 1 |
| 2.42 | -0.02 | 2 | 0 |
| 2.44 | -0.04 | 3 | 0 |
| 2.58 | -0.18 | 5 | 0 |
| 2.62 | -0.22 | 7 | 0 |
| 2.65 | -0.25 | 9 | 0 |
| 2.73 | -0.33 | 10 | 0 |
| 2.92 | -0.52 | 12 | 0 |

The test statistic, computed from Equation (25.15), is :

$$W = \sum_{i=1}^{12} r_i = 30$$

For $n = 12$ and $\alpha = 0.05$, the values $w_{0.05} = 18$ and $w_{0.975} = [(12)(13)/2] - 18 = 60$ are obtained from Table T-16. Because $18 < 30 < 60$, the null hypothesis is not rejected at the $\alpha = 0.05$ level of significance. There is insufficient statistical evidence to conclude that the median differs from 2.40. This conclusion is in agreement with the results of the sign test (Section 25.4) and the parametric Student's t -test that is also given in Section 25.4.

In these last two sections, we have considered two single-sample nonparametric methods for testing hypotheses about the population median: the sign test and the Wilcoxon signed ranks test. These are nonparametric counterparts of the

single sample t -test. There is a hierarchy in the assumptions required for these three tests. The sign test assumes only a random sample from the population. The Wilcoxon signed ranks test adds the assumption that the distribution is symmetric about the median. The t -test adds the assumption that the distribution is normal. The choice of which test to use should be based on how realistic the assumptions of each test are for the problem at hand. If the assumptions for a given test are valid, that test is more powerful for detecting deviations from H_0 than the tests lower in the hierarchy (i.e., tests with fewer restrictive assumptions).

25.6 Test of locations with two matched samples: **Sign test**

The nonparametric analysis of paired, or matched, observations is analogous to the parametric Student's t -test for paired observations (Section 15.3). With paired observations, we analyze the differences between members of the pairs, thus reducing two samples into one. Obtaining data in pairs is quite common. Some examples are:

- investigation of shipper-receiver differences.
- comparison of travel expenses as submitted by the traveller and as calculated by the auditor.
- comparison of two analytical laboratories (or instruments, technicians, etc.) by having each of them measure the same or identical items.

One nonparametric test for these situations is to compare the location parameters of the two populations using the paired observations. Let y_{1i} and y_{2i} denote the i^{th} observations from populations 1 and 2, respectively, and let $d_i = y_{1i} - y_{2i}$.

When we do not know the distribution function of d_i , when the central limit theorem does not apply, and when we cannot assume that the distribution of d_i is symmetric about its median, we can use the *sign test* to test equality of the two population medians, ζ_1 and ζ_2 . Equivalently, we test the hypothesis that the median of the differences, ζ_d , is zero, namely:

$$H_0: \zeta_d = 0 \quad (25.20)$$

The alternative hypothesis is either two-sided:

$$H_1: \zeta_d \neq 0 \quad (25.21)$$

or one-sided, stated as one of the two following statements:

$$H_1: \xi_d > 0 \tag{25.22}$$

$$H_1: \xi_d < 0 \tag{25.23}$$

As in Section 25.4, the test statistic T is the total number of plus signs for the sample $\{d_i\}$. Calculate the tail probability Q from Equation (25.11) or (25.12) and reject H_0 if $Q \leq \alpha/2$, concluding that the two populations do not have the same median.

Example 25.6. Comparison of two scales. In Example 15.1 (Section 15.3), we tested the equivalence of two scales used to weigh containers of special material by weighing each of $n = 10$ containers once on each scale. We use a two-sided sign test for the same data, with $\alpha = 0.05$.

Table 25.4 is Table 15.1 with the plus and minus signs added for the sign test.

Table 25.4. Container weights, in kg, measured by two scales

| Container | Scale 1 | Scale 2 | Difference | Sign |
|-----------|---------|---------|------------|-------------|
| 1 | 25.6 | 25.4 | -0.2 | - |
| 2 | 21.3 | 21.1 | -0.2 | - |
| 3 | 21.3 | 21.7 | 0.4 | + |
| 4 | 28.4 | 28.4 | 0.0 | 0 (ignored) |
| 5 | 29.9 | 30.0 | 0.1 | + |
| 6 | 30.0 | 29.9 | -0.1 | - |
| 7 | 23.4 | 23.1 | -0.3 | - |
| 8 | 29.5 | 29.6 | 0.1 | + |
| 9 | 27.7 | 27.5 | -0.2 | - |
| 10 | 21.3 | 21.1 | -0.2 | - |

The reduced sample size is $n = 9$. We have 3 plus signs and hence $t = 3$. The tail probability for $t = 3, n = 9$ is obtained from Equation (25. 11) and Table T-8 as $Q = 0.0020 + 0.0176 + 0.0703 + 0.1641 = 0.2540$ (rounded to 0.254). Excel’s function =BINOMDIST(3, 9, 0.5, 1) returns the value of 0.2539 (also rounded to 0.254). Because $Q > \alpha/2 = 0.025$, we do not reject H_0 , concluding that there is no statistical evidence that the two scale medians differ. This conclusion agrees with the parametric test used for the same data in Example 15.1, where the test statistic is calculated as $t = 0.896$ and $t_{0.975}(9) = 2.26$].

When a one-sided test is required, we use the same procedure but compare Q to either the α or $1 - \alpha$ quantile of the binomial distribution, as appropriate.

25.7 Test of locations with two samples: Wilcoxon matched pairs test

Wilcoxon matched pairs test The *Wilcoxon matched pairs test* also tests whether the medians of two populations are the same when the sample observations are paired. The test assumes that the distribution of the differences between the members of each pair is symmetric.

The Wilcoxon matched pairs test is the Wilcoxon signed ranks test applied to the differences between the paired observations. Let y_{1i} and y_{2i} denote the i^{th} pair of observations from populations 1 and 2, respectively. For each pair, calculate $d_i = y_{1i} - y_{2i}$ and its absolute value $|d_i|$. As for the signed ranks test, omit all $d_i = 0$ and adjust n accordingly. Rank the $|d_i|$ in ascending order and assign their average rank to all tied $|d_i|$. Let $R_i = 0$ if d_i is negative and R_i = the rank of $|d_i|$ if d_i is positive.

As in Equation (25.15), calculate $W = \sum R_i$ and compare it to $w_{\alpha/2}$ for a two-sided test and to w_α or $1 - w_\alpha$ for a one-sided test, where w_q is found in Table T-16. The null hypothesis of equality of the medians is rejected using Equation (25.16), (25.17), or (25.18), as applicable.

The test is demonstrated by the data in Table 25.5, which augments Table 25.4 by including the ranks of the differences and the associated R_i . As in Example 25.6, we use $\alpha = 0.05$.

Table 25.5. Container weights and ranks

| Container | Scale 1 | Scale 2 | Difference | Rank | R_i |
|-----------|---------|---------|------------|-----------|-----------|
| 1 | 25.6 | 25.4 | -0.2 | 3.5 | 0 |
| 2 | 21.3 | 21.1 | -0.2 | 3.5 | 0 |
| 3 | 21.3 | 21.7 | 0.4 | 9 | 9 |
| 4 | 28.4 | 28.4 | 0.0 | (omitted) | (omitted) |
| 5 | 29.9 | 30.0 | 0.1 | 7.5 | 7.5 |
| 6 | 30.0 | 29.9 | -0.1 | 6 | 0 |
| 7 | 23.4 | 23.1 | -0.3 | 1 | 0 |
| 8 | 29.5 | 29.6 | 0.1 | 7.5 | 7.5 |
| 9 | 27.7 | 27.5 | -0.2 | 3.5 | 0 |
| 10 | 21.3 | 21.1 | -0.2 | 3.5 | 0 |

From Table 25.5 we get $w = 9 + 7.5 + 7.5 = 24$. Using Equation (25.15) and Table T-16 for $n = 9$ we have $w_{0.025} = 6$ and $w_{0.975} = 45 - 6 = 39$. Since $6 < w < 39$, we have no statistical evidence that the two scales are different. This is the same conclusion we reached using the sign test and the parametric test in Example 15.1.

25.8 Test of locations; two independent samples: **Wilcoxon rank sum test**

This section presents a nonparametric test for comparing the locations of two populations based on two independent samples, one from each population. The test was developed by Wilcoxon (1945) for equal size samples and generalized by Mann and Whitney (1947) to arbitrary sample sizes. The test is referred to as the **Wilcoxon rank sum test** or as the **Mann-Whitney test**. In this text we refer to it as the **WRS test**, as used by NRC in NUREG 1505 (1998).

Wilcoxon rank sum test

Mann-Whitney test

WRS test

- A condensed version of the WRS test may be found in Wilcoxon and Wilcox (1964, p.7).

Let $\{ y_{11}, y_{12}, \dots, y_{1n} \}$ be a sample of size n from a population with median ζ_1 and let $\{ y_{21}, y_{22}, \dots, y_{2m} \}$ be an independent sample of size m from a second population with median ζ_2 . We wish to test the hypothesis:

$$H_0: \zeta_1 = \zeta_2 \quad (25.24)$$

The alternative hypothesis H_1 may be a one- or a two-sided alternative.

The WRS test assumes that the two population distributions are the same except for a possible difference in their medians. If the distributions are also symmetric, then the test for equality of medians is also a test of equality of means.

The WRS test combines the $m + n$ observations from the two samples and ranks the combined sample from 1 to $m + n$. If any sample values are tied, each is assigned the average of the ranks that would otherwise have been assigned. Let $R(y_{1i})$ denote the rank assigned to y_{1i} , $i = 1, \dots, n$, and let $R(y_{2j})$ denote the rank assigned to y_{2j} , $j = 1, \dots, m$.

The test statistic for testing H_0 is the sum of the ranks from the first sample:

$$W = \sum_{i=1}^n R(y_{1i}) \quad (25.25)$$

- The test statistic could just as well have been defined as the sum of the ranks from the second sample, because the labeling of the populations is arbitrary.

- Even though the same symbol is used for the test statistic for the WRS test as for the Wilcoxon signed ranks test (Equation (25.15)), there should be no confusion.

Critical values $w_q(n, m)$ of the W statistic are given in Table T-17 of the appendix for $n \leq 20$ and $m \leq 20$ for the lower quantiles $q = 0.001, 0.005, 0.01, 0.025, 0.05,$ and 0.10 . Critical values for the corresponding upper quantiles are given by:

$$w_q(n, m) = n(n+m+1) - w_{1-q}(n, m) \quad (25.26)$$

When either $m > 20$ or $n > 20$, the q^{th} quantile of the W statistic is approximated by:

$$w_q(m, n) = n(n+m+1)/2 + z_q \sqrt{nm(n+m+1)/12} \quad (25.27)$$

where z_q is the q^{th} quantile of the standard normal distribution, Table T-1 of the appendix.

For $H_0: \xi_1 = \xi_2$, the critical values for the various alternative hypothesis are given by:

$$\text{For } H_1: \xi_1 \neq \xi_2, \text{ reject } H_0 \text{ if } w > w_{1-\alpha/2}(n, m) \text{ or if } w < w_{\alpha/2}(n, m) \quad (25.28)$$

$$\text{For } H_1: \xi_1 > \xi_2, \text{ reject } H_0 \text{ if } w > w_{1-\alpha}(n, m) \quad (25.29)$$

$$\text{For } H_1: \xi_1 < \xi_2, \text{ reject } H_0 \text{ if } w < w_{\alpha}(n, m) \quad (25.30)$$

Example 25.7. Comparison of two batches (Bowen and Bennett (1988), p. 497)). A facility producing UO_2 powder wishes to check whether the percent uranium factor for two batches of powder is the same. Eight samples from each of the two batches are drawn and the percent uranium determined. The data are displayed in Table 25.6. Use a two-sided WRS test with $\alpha = 0.05$.

Table 25.6. Percent uranium in UO₂

| Batch 1 | Batch 2 |
|---------|---------|
| 86.4 | 86.2 |
| 86.8 | 85.8 |
| 87.5 | 86.1 |
| 88.2 | 84.8 |
| 86.0 | 87.1 |
| 86.9 | 86.5 |
| 87.7 | 85.7 |
| 86.8 | 86.7 |

The validity of the assumptions underlying the *WRS* test should be checked. The samples from each batch are assumed to be randomly collected. Application of the runs test (Section 25.3) to each sample does not find any statistical evidence of non-randomness.

The *WRS* test calls for ranking the observations in the combined sample. This process is easier if the individual samples are ordered first, as shown in Table 25.7, along with the combined ranks.

Table 25.7. Ranks of combined samples from Batches 1 and 2

| Batch 1 | Combined rank | Batch 2 | Combined rank |
|---------|---------------|---------|---------------|
| 86.0 | 4 | 84.8 | 1 |
| 86.4 | 7 | 85.7 | 2 |
| 86.8 | 10.5 | 85.8 | 3 |
| 86.8 | 10.5 | 86.1 | 5 |
| 86.9 | 12 | 86.2 | 6 |
| 87.5 | 14 | 86.5 | 8 |
| 87.7 | 15 | 86.7 | 9 |
| 88.2 | 16 | 87.1 | 13 |

The test statistic is calculated from Equation (25.25), yielding:

$$W = \sum_{i=1}^8 R(y_{i1}) = 89$$

For $n = m = 8$ and $\alpha/2 = 0.025$, the values $w_{0.025} = 50$ is obtained from Table T-17. Using Equation (25.26), $w_{0.975} = 8(8 + 8 + 1) - 50 = 86$. Because $89 > 86$, we reject H_0 , concluding that the percent uranium for the two batches are different.

The parametric counterpart to the *WRS* test is the two-sample *t*-test of Section 15.5, where the data are assumed to be drawn from two normal distributions with equal variances. Assuming normality for the data in Example 25.7, we have the following results, using Excel's data analysis routine:

Table 25.8. Parametric test for Example 25.7

| | Batch 1 | Batch 2 |
|---|---------|---------|
| Mean | 87.04 | 86.11 |
| Variance | 0.52 | 0.50 |
| Observations | 8 | 8 |
| Pooled variance | 0.506 | |
| Hypothesized mean difference | 0 | |
| Degrees of freedom | 14 | |
| <i>t</i>-statistic | 2.60 | |
| <i>t</i> critical, two-tail ($\alpha = 0.05$) | 2.14 | |

Since $2.60 > 2.14$ we have statistical evidence that the two batches came from populations with different means. This conclusion is the same that is reached using the *WRS* test.

An argument can be made for always using the *WRS* test for comparing two samples. First, although the *t*-test is more powerful than the *WRS* test if the population distributions are normal, the difference in power is small. However, because the *t*-test is not valid for non-normal distributions, the *WRS* test is more powerful than the *t*-test for many non-normal distributions. Therefore, the *WRS* test is preferred if the population distributions are unknown or if there is any doubt about the normality assumption and it loses little if the distributions are actually normal.

25.9 Test of locations with several samples:

The median test

This section presents a nonparametric test for comparing the locations of several populations. The two most common nonparametric tests for comparing locations are the median test, presented in this section, and the Kruskal-Wallis

test, presented in the next section. The parametric test that corresponds to these tests is the one-way ANOVA.

median test The **median test** investigates whether k independent samples come from populations having the same median. If the populations are symmetric, testing equality of the medians is equivalent to testing equality of the means. The test assumes that the samples are collected randomly and independently and that the observations are at least on an ordinal scale.

Let ζ_i denote the median of the i^{th} population and let n_i be the number of observations in the i^{th} sample, $i = 1, \dots, k$. The null hypothesis is:

$$H_0: \zeta_1 = \zeta_2 = \dots = \zeta_k \tag{25.31}$$

The alternative hypothesis may be written as:

$$H_1: \zeta_i \neq \zeta_j, \text{ some } i, j \tag{25.32}$$

To use the median test, first calculate the grand median M of the combined samples. Let O_{1i} be the number of observations in the i^{th} sample that exceed M and let O_{2i} be the number less than or equal to M . Arrange the O_{1i} and O_{2i} , $i = 1, 2, \dots, k$, into a $2 \times k$ contingency table as shown in Table 25.9, where a denotes the total counts of the observations larger than M and b , the total counts of the observations that are equal or smaller than M .

Table 25.9. Contingency table for the median test

| | Sample | | | | Total |
|-----------------------|----------|----------|-----|----------|---------|
| | 1 | 2 | ... | k | |
| Observations $> M$ | O_{11} | O_{12} | ... | O_{1k} | A |
| Observations $\leq M$ | O_{21} | O_{22} | ... | O_{2k} | B |
| Total | n_1 | n_2 | ... | n_k | $a + b$ |

The test statistic is given by:

$$\chi^2 = \frac{(a + b)^2}{ab} \sum_{i=1}^k \frac{O_{1i}^2}{n_i} - \frac{a(a + b)}{b} \tag{25.33}$$

Alternatively, we can use either Equation (12.5) or (12.6) for the $r \times k$ contingency table analysis of Chapter 12. The numerical result obtained from

the equations in Chapter 12 are identical to that of Equation (25.44), but the calculations from equation (25.44) are usually simpler.

The test statistic χ^2 is approximately distributed as a chi-square random variable with $\nu = k - 1$ degrees of freedom. Hence we reject H_0 if $\chi^2 > \chi_{1-\alpha}^2(k-1)$, where $\chi_{1-\alpha}^2(k-1)$ is obtained from Table T- 2 of the appendix.

The chi-square approximation may not be very accurate if some of the n_i are small. Conover ((1980), p. 172) suggests that, in general, the approximation may not be satisfactory if more than 20% of the n_i 's are less than 10 or if any of the n_i 's are less than 2. An exception to this rule can be made for larger values of k ; if most of the n_i 's are about the same, then values of n_i as small as 2 are allowed.

Example 25.8. Comparison of four production lines (Bowen and Bennett, 1988, p. 502). A facility has four production lines that each create a liquid waste stream. All four of the waste streams go to the same storage area, where they are sampled and analyzed (uranium concentration in ppm) for accountability purposes. The facility must test whether the four waste streams are entering the storage area with the same uranium concentration. Random samples from the four waste streams are collected and analyzed for uranium concentrations. The results are displayed in Table 25.11. Use the median test to test the null hypothesis that there are no differences between the four production lines, with $\alpha = 0.10$.

Table 25.10. Uranium concentration (ppm)

| Line 1 | Line 2 | Line 3 | Line 4 |
|--------|--------|--------|--------|
| 13 | 34 | 17 | 33 |
| 23 | 16 | 26 | 30 |
| 23 | 20 | 11 | 23 |
| 14 | 35 | 13 | 24 |
| 25 | 17 | 24 | 32 |
| 22 | 30 | 22 | 20 |
| 35 | 27 | 8 | 17 |
| 13 | 28 | 20 | 32 |
| 24 | | 23 | |
| 20 | | | |

The grand median of the combined samples is 23. Accordingly, the following 2 x 4 contingency table (Table 25.11) is constructed in the template of Table 25.9.

Table 25.11. Contingency table for Example 25.8

| | Line | | | | Total |
|--------------|------|---|---|---|-------|
| | 1 | 2 | 3 | 4 | |
| > 23.0 | 3 | 5 | 2 | 5 | 15 |
| < 23.0 | 7 | 3 | 7 | 3 | 20 |
| Total | 10 | 8 | 9 | 8 | 35 |

The test statistic, calculated from Equation 25.33, is:

$$\chi^2 = \frac{(35)^2}{(15)(20)} \left[\frac{3^2}{10} + \frac{5^2}{8} + \frac{2^2}{9} + \frac{5^2}{8} \right] - \frac{(35)(15)}{20} = 31.01 - 26.25 = 4.76$$

From Table T-2, the critical value for $\nu = 3$ and $\alpha = 0.10$ is $\chi^2_{0.90}(3) = 6.25$. Because $4.76 < 6.25$, H_0 is not rejected. There is no significant evidence that the median uranium concentrations for the four lines are different.

The parametric counterpart of the median test is a one-way ANOVA. Table 25.12 displays Excel’s ANOVA for the data in Example 25.8, where $\alpha = 0.10$. In that analysis the calculated F statistic is 2.94, and the critical value is $f_{0.10}(3,31) = 2.27$. Because $2.94 > 2.27$, H_0 is rejected and we conclude that the mean uranium concentrations are not all equal. This result is not that surprising since the parametric test, operating under more stringent assumptions, is more powerful than the median test in detecting differences between the populations.

Table 25.12. Excel’s output for Example 25.8

| SUMMARY | | | | |
|---------|-------|-----|---------|----------|
| Groups | Count | Sum | Average | Variance |
| Line 1 | 10 | 212 | 21.2 | 45.29 |
| Line 2 | 8 | 207 | 25.9 | 54.70 |
| Line 3 | 9 | 164 | 18.2 | 39.94 |
| Line 4 | 8 | 211 | 26.4 | 37.98 |

| ANOVA | | | | | |
|---------------------|---------|----|--------|------|--------|
| Source of Variation | SS | df | MS | F | F crit |
| Between Groups | 391.64 | 3 | 130.55 | 2.94 | 2.27 |
| Within Groups | 1375.91 | 31 | 44.38 | | |
| Total | 1767.54 | 34 | | | |

25.10 Test of locations with several samples: The Kruskal-Wallis test

Kruskal-Wallis test Like the median test, the **Kruskal-Wallis test** is used to compare the locations of several populations. As for the median test, we assume that we have k mutually independent samples from k populations with data on, at least, an ordinal scale. An additional assumption for the Kruskal-Wallis test is that the population distributions are identical, except possibly for their locations.

➤ This assumption implies that the population variances are identical.

Because equality of the medians is equivalent to equality of the means, the null and alternative hypotheses can be written as:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad (25.34)$$

$$H_1: \mu_i \neq \mu_j, \text{ some } i, j \quad (25.35)$$

where μ_i is the mean of the i^{th} population, $i = 1, 2, \dots, k$.

Denote the sample from the i^{th} population by $y_{i1}, y_{i2}, \dots, y_{in_i}$, where n_i is the sample size, $i = 1, 2, \dots, k$. Let $n = \sum n_i$ be the total number of observations for the k samples.

Rank the combined samples (smallest to largest) from 1 to n . If several observations are tied, assign to each the average of the ranks that would otherwise have been assigned. Let $R(y_{ij})$ denote the rank assigned to y_{ij} , and let R_i be the sum of the ranks assigned to the i^{th} sample; thus,

$$R_i = \sum_{j=1}^{n_i} R(y_{ij}), i = 1, \dots, k \quad (25.36)$$

The Kruskal-Wallis statistic for testing equality of the means is:

$$H = \frac{1}{c} \left[\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{n(n+1)^2}{4} \right] \quad (25.37)$$

where

$$c = \frac{n(n+1)}{12}, \text{ when no ties are present} \quad (25.38)$$

$$c = \frac{1}{n-1} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} R(Y_{ij})^2 - \frac{n(n+1)^2}{4} \right], \quad (25.39)$$

when ties are present

The H statistic is distributed approximately as a chi-square variable with $\nu = k - 1$ degrees of freedom. The null hypothesis of equal means is rejected the at α level of significance if $H > \chi^2_{1-\alpha}(k - 1)$, and $\chi^2_{1-\alpha}(k - 1)$ is the $(1-\alpha)$ quantile of the chi-square distribution of Table T-2 of the appendix.

If we reject H_0 , all we can conclude is that at least two population means are different. To identify which specific means are different, we run a multiple comparison test that compares all possible pairs of means. For any two populations, i and j , we conclude that their means are different if the absolute value of the difference $R_i/n_i - R_j/n_j$ between their average ranks exceeds an appropriate critical value. Specifically, we reject the hypothesis that $\mu_i = \mu_j$ if

$$\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right| > t_{1-\alpha/2}(n-k) \sqrt{c \frac{n-1-H}{n-k}} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad (25.40)$$

where H is given by Equation (25.37) and c is given by Equation (25.49) if the data contains no ties or Equation (25.39) otherwise. The factor $t_{1-\alpha/2}(n-k)$ is obtained from Table T-3 of the appendix for $\nu = n - k$ degrees of freedom. The test is repeated for all pairs of populations.

Example 25.9. Comparison of four production lines (continued). In Example 25.8, we tested the equality of uranium concentration medians of four production lines at the $\alpha = 0.10$ level of significance using the median test. Assume that the four distributions may differ only in their locations so that we can use the Kruskal-Wallis test to compare the means.

The data are shown in Table 25.10. They are copied in Table 25.13, along with their ranks in the combined sample of $n = 35$ observations, as required by the Kruskal-Wallis test. The bottom row of the table shows R_i , the sum of the ranks for the i^{th} sample.

Table 25.14. Data ranked for Example 25.9

| Line 1 | | Line 2 | | Line 3 | | Line 4 | |
|----------|-------|----------|-------|----------|-------|----------|-------|
| y_{1j} | Rank | y_{2j} | Rank | y_{3j} | Rank | y_{4j} | Rank |
| 13 | 4 | 34 | 7 | 17 | 1 | 33 | 9 |
| 23 | 4 | 16 | 9 | 26 | 2 | 30 | 12.5 |
| 23 | 6 | 20 | 12.5 | 11 | 4 | 23 | 18.5 |
| 14 | 12.5 | 35 | 26 | 13 | 9 | 24 | 22 |
| 25 | 15.5 | 17 | 27 | 24 | 12.5 | 32 | 28.5 |
| 22 | 18.5 | 30 | 28.5 | 22 | 15.5 | 20 | 30.5 |
| 35 | 18.5 | 27 | 33 | 8 | 18.5 | 17 | 30.5 |
| 13 | 22 | 28 | 34.5 | 20 | 22 | 32 | 32 |
| 24 | 24 | | | 23 | 25 | | |
| 20 | 34.5 | | | | | | |
| $R_1 =$ | 159.5 | $R_2 =$ | 177.5 | $R_3 =$ | 109.5 | $R_4 =$ | 183.5 |

Since some of the ranks are tied, we calculate c using Equation (25.50). We have:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} [R(y_{ij})]^2 = 4^2 + 4^2 + \dots + 32^2 = 14892$$

$$c = \frac{1}{34} \left[14892 - \frac{35(36)^2}{4} \right] = 104.47$$

Next we calculate $\sum_{i=1}^4 \frac{R_i^2}{n_i} = \frac{159.5^2}{10} + \frac{159.5^2}{8} + \frac{159.5^2}{9} + \frac{159.5^2}{8} = 12023.59$

Finally, using Equation (25.37), Kruskal-Wallis test statistic is calculated as:

$$H = \frac{1}{104.47} \left[12024 - 35(36)^2/4 \right] = 6.55$$

The critical value for the test is $\chi^2_{0.90}(3) = 6.25$. Because $6.55 > 6.25$, H_0 is rejected at the $\alpha = 0.10$ level of significance. This conclusion agrees with the parametric ANOVA shown in Table 25.13, but not with the conclusion of the median test for Example 25.8. These divergent results are due to the greater sensitivity of the Kruskal-Wallis and the parametric tests to differences in the distributions of the different populations.

Because the Kruskal-Wallis test rejected the hypothesis of equal population means, we continue with the multiple comparison test to identify which pairs of means are different. From Equation (25.40), calculate

$$t_{1-\alpha/2}(n-k) \sqrt{C \frac{n-1-H}{n-k}} = (1.70) \sqrt{104.47 \frac{35-1-6.55}{35-4}} = 16.35$$

Next, apply Equation (25.40) to all pairs of differences as shown in Table 25.14.

Table 25.14. Multiple comparisons for Example 25.9

| Population
<i>i</i> <i>j</i> | | $\left \frac{R_i}{n_i} - \frac{R_j}{n_j} \right $ | $16.35 \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ | Significant at
$\alpha = 0.10$ |
|---------------------------------|---|--|--|-----------------------------------|
| 1 | 2 | 6.24 | 7.76 | No |
| 1 | 3 | 3.78 | 7.51 | No |
| 1 | 4 | 6.99 | 7.76 | No |
| 2 | 3 | 10.02 | 7.94 | Yes |
| 2 | 4 | 0.75 | 8.18 | No |
| 3 | 4 | 10.77 | 7.94 | Yes |

These results indicate that the mean uranium concentration for Line 3 is significantly different (lower) than the mean concentrations for Lines 2 and 4.

The same results are displayed schematically as in Duncan’s multiple range test of Section 16.8. The number listed for each line is the average rank for the corresponding sample. Means that are not underscored by the same line are significantly different.

| Line 3 | Line 1 | Line 2 | Line 4 |
|--------|--------|--------|--------|
| 12.17 | 15.95 | 22.19 | 22.94 |

Both the median and the Kruskal-Wallis tests are used for comparing the locations of several populations. The median test requires only that the samples be independent but the Kruskal-Wallis test also requires that the populations all have identical distribution functions, except for possibly different means. If this more stringent assumption holds, the Kruskal-Wallis test is more powerful for detecting population differences than the median test because ranking the data retains more information than does dichotomizing. However, if the population distributions differ by more than location, then the Kruskal-Wallis test is not applicable and the median test should be used.

25.11 Test of location for related samples: **The rank ANOVA test**

Tests for comparing the locations of several populations using related samples are extensions of two-population comparisons when the observations are matched as pairs (Section 25.7). The parametric equivalent of these tests is the two-way ANOVA for a randomized complete block design (Section 17.4).

The analysis of randomized complete block designs compare the means resulting from $k \geq 2$ different treatments, where the observations are arranged in blocks. The blocks for these designs are groups of k experimental units where the units within a group are similar, but significant differences among the groups may be the results of the different experimental treatments. If b batches of homogeneous material are split into k portions and each portion is analyzed by each of k laboratories, the k laboratories are considered treatment and the b batches are considered blocks.

The purpose of the statistical analysis of the response data is to determine whether the k treatments are alike in their locations, such as medians or means. Denote those locations by ζ_i , $i = 1, \dots, k$. The null and the alternative hypothesis are the same as stated for the median test (and also applicable to the Kruskal-Wallis test) in Equation (25.31) and (25.32):

$$H_0: \zeta_1 = \zeta_2 = \dots = \zeta_k = \zeta$$

and

$$H_1: \zeta_i \neq \zeta_j, \text{ some } i, j$$

Rank ANOVA test

In this section we provide the **Rank ANOVA test**, which is considered to be the most powerful nonparametric (sensitive to departure from H_0) test for many applications. It should be used when the distributions we investigate are identical, with possible different locations. However, if we have a reason to believe that the underlying distribution is rather flat, the Friedman Test of Section 25.12 would be a better choice.

The Rank ANOVA test is identical to the randomized block design ANOVA of Sections 17.3 - 17.5, except that the observations are replaced by their rank, where the ranking of the data is made across all $b \times k$ data cells. Tied observations are assigned their average rank.

The ANOVA calculations follow the identical steps that are outlined in Section 17.5. The test statistic is the F statistic with $v_1 = k - 1$ degrees of freedom in the numerator and $v_2 = (k - 1)(b - 1)$ degrees of freedom in the denominator. If the calculated value of F exceeds the critical value $f_{1-\alpha}(v_1, v_2)$, H_0 is rejected and we conclude that not all population locations are the same.

The use of the rank ANOVA test is demonstrated in Example 25.11.

Example 25.10. Percent plutonium by weight (Bowen and Bennett, p. 612). Each of three laboratories was given a sample of PuO_2 from each of 10 containers. The laboratories measured the plutonium by weight of each sample. In turn, each weight was replaced by its rank. We test whether the population locations are equal, using $\alpha = 0.05$ level of significance.

Table 25.15 shows the weights and the ranks of the data by block (container) and treatment (laboratory)

Table 25.15. Data and ranks of plutonium data

| Container | Lab 1 | Rank | Lab 2 | Rank | Lab 3 | Rank |
|-----------|-------|------|-------|------|-------|------|
| 1 | 87.67 | 5 | 87.88 | 19 | 87.72 | 8.5 |
| 2 | 87.64 | 3 | 87.99 | 25 | 87.80 | 14 |
| 3 | 88.03 | 26.5 | 87.98 | 24 | 87.97 | 23 |
| 4 | 87.50 | 1 | 87.84 | 17 | 87.65 | 4 |
| 5 | 87.86 | 18 | 87.77 | 12 | 87.75 | 10 |
| 6 | 87.71 | 7 | 87.93 | 20.5 | 87.76 | 11 |
| 7 | 87.63 | 2 | 87.83 | 15.5 | 87.79 | 13 |
| 8 | 87.83 | 15.5 | 87.94 | 22 | 87.93 | 20.5 |
| 9 | 88.04 | 28 | 88.06 | 30 | 88.05 | 29 |
| 10 | 87.72 | 8.5 | 88.03 | 26.5 | 87.68 | 6 |

We could use all the steps given in Section 17.5 (ANOVA for a two-way factorial design without replication) to conduct the required analysis. We elected instead to use Excel's "ANOVA: Two Factors Without Replication," routine in the Data Analysis module. In that routine, we identified the location of the data in the "Input Range" box, checked the "Label" box, provided the level of significance as 0.05 in the "Alpha" box, and gave the address for the output in the "Output Range." Excel's analysis (modified to fit this book style) is shown in Table 25.16.

Table 25.16. Excel's output for Example 25.11

| SUMMARY | | | | |
|---------|-------|-------|---------|----------|
| Groups | Count | Sum | Average | Variance |
| Lab 1 | 10 | 114.5 | 11.45 | 99.97 |
| Lab 2 | 10 | 211.5 | 21.15 | 29.84 |
| Lab 3 | 10 | 139 | 13.9 | 63.16 |

| ANOVA | | | | | |
|---------------------|----------|----|---------|------|--------|
| Source of Variation | SS | df | MS | F | F crit |
| Rows | 1236.833 | 9 | 137.426 | | |
| Columns | 508.850 | 2 | 254.425 | 9.16 | 3.55 |
| Error | 499.817 | 18 | 27.768 | | |
| Total | 2245.500 | 29 | | | |

The critical value for this test is $f_{0.95}(9, 18) = 3.55$. Because $9.16 > 3.55$, we have statistical evidence that not all population locations are the same.

25.12 Test of location for related samples: The Friedman test

Friedman test As indicated earlier, there are several tests for testing equality of population locations for related samples. This section introduces the **Friedman test** which is an extension of the matched pairs of the sign test to multiple populations.

The data structure to which the Friedman test is applied is the randomized complete block design. In this design it is assumed that the k populations under consideration have the same distribution, with possibly different locations.

The null hypothesis of the test is the same as in the rank ANOVA test, that is, the population locations of the treatments are the same. The alternative hypothesis is that at least two treatments have different locations.

In the setup of the Friedman test, the observations are arranged in k columns (corresponding to k treatments) and b rows (corresponding to b blocks). Let y_{ij} denote the observation from the j^{th} block for the i^{th} treatment. Let $R(y_{ij})$ denote the rank of y_{ij} within the j^{th} block, and where tied observations receive their average rank. We define three quantities:

$$R_i = \sum_{j=1}^b R(y_{ij}) \quad (25.41)$$

$$A = \sum_{i=1}^k \sum_{j=1}^b [R(y_{ij})]^2 \quad (25.42)$$

$$B = \frac{1}{b} \sum_{i=1}^k R_i^2 \quad (25.43)$$

The test statistic is given by:

$$F = \frac{(b-1)[B - bk(k+1)^2 / 4]}{A - B} \quad (25.44)$$

The *test statistic* is distributed approximately as an F variable with $v_1 = k - 1$ degrees of freedom in the numerator and $v_2 = (b - 1)(k - 1)$ degrees of freedom in the denominator. If the calculated F exceeds the critical value $f_{1-\alpha}(v_1, v_2)$, the null hypothesis of equal population locations is rejected.

- The approximation of the test statistic in Equation (25.44) to the F -distribution improves as the number of blocks b increases.

We apply the Friedman test to the data in Example 25.10 where $\alpha = 0.05$. The observations and their within-block ranking are shown in Table 25.17.

Table 25.18. Data and ranks of plutonium data for the Friedman test

| Container | Lab 1 | Rank | Lab 2 | Rank | Lab 3 | Rank |
|-----------|-------|------|-------|------|-------|------|
| 1 | 87.67 | 1 | 87.88 | 3 | 87.72 | 2 |
| 2 | 87.64 | 1 | 87.99 | 3 | 87.80 | 2 |
| 3 | 88.03 | 3 | 87.98 | 2 | 87.97 | 1 |
| 4 | 87.50 | 1 | 87.84 | 3 | 87.65 | 2 |
| 5 | 87.86 | 3 | 87.77 | 2 | 87.75 | 1 |
| 6 | 87.71 | 1 | 87.93 | 3 | 87.76 | 2 |
| 7 | 87.63 | 1 | 87.83 | 3 | 87.79 | 2 |
| 8 | 87.83 | 1 | 87.94 | 3 | 87.93 | 2 |
| 9 | 88.04 | 1 | 88.06 | 3 | 88.05 | 2 |
| 10 | 87.72 | 2 | 88.03 | 3 | 87.68 | 1 |

The calculations are:

$$R_1 = \sum_{j=1}^{10} R(y_{1j}) = 15, \quad R_2 = \sum_{j=1}^{10} R(y_{2j}) = 28, \quad R_3 = \sum_{j=1}^{10} R(y_{3j}) = 17$$

$$A = \sum_{i=1}^3 \sum_{j=1}^{10} R(y_{ij})^2 = 140$$

$$B = \frac{1}{10} \sum_{i=1}^3 R_i^2 = \frac{15^2 + 28^2 + 17^2}{10} = 129.8$$

The test statistic is calculated from Equation (25.44) as:

$$f = \frac{(10-1)[129.8 - (10)(3)(3+1) / 4]}{140 - 129.8} = 8.65$$

The critical value for the test, obtained by interpolation from Table T-4 of the appendix or by a call to Excel, is $f_{0.95}(2, 18) = 3.55$. Because $8.65 > 3.55$ we

have statistical evidence that not all population locations are equal. These results agree with the parametric ANOVA of Table 25.16.

Quade test As stated earlier, there are other nonparametric tests for testing equality of population locations. The **Quade test**, for example, is the chief competitor to the Friedman test. We elected to present the Friedman test because it is more commonly used. The interested reader may refer to Conover (1980, p. 296) or to Bowen and Bennett (1988, p.610) for details about the Quade test.

25.13 Test of variances with two samples: Squared ranks test

Except for the runs test for randomness, all of the nonparametric tests presented so far have been tests of location. We now consider tests of variability, as measured by the variance.

squared ranks test The variances of two populations can be compared with a nonparametric test called the **squared ranks test**, developed by Conover and Iman (1978). We assume that the two populations have identical distributions, either continuous or discrete, except possibly for different means and variances. For example, the populations might be normally distributed. Denote their means by μ_1 and μ_2 and their variances by σ_1^2 and σ_2^2 .

Let $\{y_{11}, y_{12}, \dots, y_{1n}\}$ and $\{y_{21}, y_{22}, \dots, y_{2m}\}$ be independent random samples of size n and m , respectively, from the two populations. We test the null hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (25.45)$$

against a two- or one-sided alternative hypothesis.

The **squared ranks test** of H_0 is performed as follows:

$$u_i = |y_{1i} - \mu_1|, \quad i = 1, 2, \dots, n \quad (25.46)$$

$$v_j = |y_{1j} - \mu_2|, \quad j = 1, 2, \dots, m \quad (25.47)$$

If μ_1 and μ_2 are unknown, the test is still approximately valid if we use the sample means \bar{y}_1 and \bar{y}_2 instead. Rank the $n + m$ observations in the combined sample of u_i 's and v_j 's. If any values of u_i or v_j are tied, assign to each the

average of the ranks that would have been assigned had there been no ties. Denote the resulting ranks by $R(u_i)$ and $R(v_j)$.

If there are no ties, the test statistic is the sum of the squared ranks of the u_i , denoted by T_1 :

$$T_1 = \sum_{i=1}^n [R(u_i)]^2 \quad (25.48)$$

If there are ties, the test statistic is:

$$T_1^* = \frac{T_1 - n\bar{R}^2}{\sqrt{\frac{nm}{(n+m)(n+m-1)} \sum_{i=1}^{n+m} R_i^4 - \frac{nm}{n+m-1} \bar{R}^2}} \quad (25.49)$$

where:

$$\bar{R}^2 = \frac{1}{n+m} \left\{ \sum_{i=1}^n R(u_i)^2 + \sum_{j=1}^m R(v_j)^2 \right\} \quad (25.50)$$

and

$$\sum_{k=1}^{n+m} R_k^4 = \sum_{i=1}^n R(u_i)^4 + \sum_{j=1}^m [R(v_j)]^4 \quad (25.51)$$

The quantiles for the squared rank test statistics are obtained as follows:

- When $n \leq 10$ and $m \leq 10$ and when no ties are present, exact quantiles $w_q = w_q(n, m)$, of the distribution of T_1 are given in Table T-18 of the appendix.
- When $n \leq 10$ and $m \leq 10$ and when ties are present, the test statistic T_1^* is distributed approximately as a standard normal distribution, with quantiles given in Table T-1.
- When either $n > 10$ or $m > 10$, the q^{th} quantile w_q of T_1 is as approximated by

$$w_q = \frac{n(m+n+1)(2n+2m+1)}{6} + z_q \sqrt{\frac{nm(n+m+1)(2n+2m+1)(8n+8m+11)}{180}} \quad (25.52)$$

where z_q is the q^{th} quantile of the normal distribution (Table T-1).

For $H_0: \sigma_1^2 = \sigma_2^2$, the critical values for the various alternative hypotheses are given by:

$$\begin{aligned} \text{For } H_1: \sigma_1^2 \neq \sigma_2^2, \text{ reject } H_0 \text{ if } T_1 \text{ (or } T_1^*) < w_{\alpha/2} \\ \text{or if } T_1 \text{ (or } T_1^*) > w_{1-\alpha/2} \end{aligned} \tag{25.53}$$

$$\text{For } H_1: \sigma_1^2 > \sigma_2^2, \text{ reject } H_0 \text{ if } T_1 \text{ (or } T_1^*) > w_{1-\alpha} \tag{25.54}$$

$$\text{For } H_1: \sigma_1^2 < \sigma_2^2, \text{ reject } H_0 \text{ if } T_1 \text{ (or } T_1^*) < w_{\alpha} \tag{25.55}$$

Example 25.8. (continued). Comparison of percent of uranium in two batches. We continue with a check on the assumptions of the WRS test, which requires that the population distributions be the same except for a possible location difference. We apply the squared ranks test to the data in Example 25.7 to test for equality of the variances, using $\alpha = 0.05$.

Because the population means are unknown, the sample estimates $\bar{y}_1 = 87.04$ and $\bar{y}_2 = 86.11$ are used. The absolute values of the deviations and their ranks in the combined sample are displayed in Table 25.18.

Table 25.18. Data sorted and ranked for Example 25.8

| Batch 1
(y_{1i}) | $u_i = y_{1i} - \bar{y}_1 $ | $R(u_i)$ | Batch 2
(y_{2i}) | $v_i = y_{2i} - \bar{y}_2 $ | $R(v_i)$ |
|-------------------------|------------------------------|----------|-------------------------|------------------------------|----------|
| 86.0 | 1.04 | 14 | 84.8 | 1.31 | 16 |
| 86.4 | 0.64 | 11 | 85.7 | 0.41 | 8 |
| 86.8 | 0.24 | 4.5 | 85.8 | 0.31 | 6 |
| 86.8 | 0.24 | 4.5 | 86.1 | 0.01 | 1 |
| 86.9 | 0.14 | 3 | 86.2 | 0.09 | 2 |
| 87.5 | 0.46 | 9 | 86.5 | 0.39 | 7 |
| 87.7 | 0.66 | 12 | 86.7 | 0.59 | 10 |
| 88.2 | 1.16 | 15 | 87.1 | 0.99 | 13 |

From Table 25.18 we calculate:

$$\sum_{i=1}^8 [R(u_i)]^2 = 816.5, \quad \sum_{i=1}^8 [R(u_i)]^4 = 131880$$

$$\sum_{i=1}^8 [R(v_i)]^4 = 679.0, \quad \sum_{i=1}^8 [R(v_i)]^4 = 111907$$

Because there are ties in the data, Equations (25.49), (25.50), and (25.51) are used.

$$\text{From Equation (25.50), we have } \bar{R}^2 = \frac{1}{8+8} 816.5 + 679.0 = 93.47$$

$$\text{From Equation (25.51), we have } \sum_{k=1}^{16} R_k^4 = 131880 + 111907 = 243787$$

From Equation (25.49), we have:

$$T_1^* = \frac{816.5 - 8(93.47)}{\sqrt{\frac{(8)(8)}{(8+8)(8+8-1)}(243787) - \frac{(8)(8)}{8+8-1}(93.47)^2}} = 0.41$$

From Table T-1, we obtain $w_{0.025} = -1.960$ and $w_{0.975} = 1.960$. Because $-1.960 < 0.41 < 1.960$, H_0 is not rejected. There is no statistical evidence that the variances of the two populations are different.

The parametric counterpart to the squared ranks test is the F -test, described in Section 14.4. For comparison, we can perform the F -test for Example 25.8. The two sample variances, listed in Table 25.8, are 0.52 and 0.50 with 7 degrees of freedom each. The test statistic, which is the ratio of the larger to the smaller variance, is $f = 0.52/0.50 = 1.04$. The critical value for this test is $f_{0.95}(7,7) = 4.99$. Since $1.04 < 4.99$, we have no statistical evidence that the variances of the two populations are different.

If the two samples are actually from normal distributions, the F -test would be the test statistic of choice, as the test is quite sensitive to variance differences. However, the F -test is also sensitive to departures from normality, and a statistically significant result may be due to non-normality and not to variance differences. Thus, the squared ranks test is preferred if normality is in doubt.

25.14 Test of variances with several samples: **The k -sample squared ranks test**

k-sample squared ranks test The two-sample squared ranks test can be extended to the *k*-sample squared ranks test to compare the variances of k

populations. As for the two-sample test, we assume that the k populations have identical distributions, either continuous or discrete, except possibly for different means and variances. Denote the i^{th} population mean and variance by μ_i and σ_i^2 , respectively, $i = 1, 2, \dots, k$.

The null and alternative hypotheses for testing the equality of the k variances are:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \tag{25.56}$$

$$H_1: \sigma_i^2 \neq \sigma_j^2, \text{ some } i, j \tag{25.57}$$

Denote a sample of n_i observations from the i^{th} population by $y_{i1}, y_{i2}, \dots, y_{in_i}$. From each observation y_{ij} , subtract its population mean μ_i , or its sample mean \bar{y}_i if μ_i is unknown. Combine all k samples and rank the absolute differences $|y_{ij} - \mu_i|$ or $|y_{ij} - \bar{y}_i|$ from smallest to largest, assigning average ranks in case of ties. Denote the rank of $|y_{ij} - \mu_i|$ (or $|y_{ij} - \bar{y}_i|$) by $R(Y_{ij})$. The test statistic is given by:

$$T = \frac{1}{D^2} \left[\sum_{i=1}^k \frac{S_i^2}{n_i} - n(\bar{S})^2 \right] \tag{25.58}$$

where:

n_i = the number of observations in the i^{th} sample

$$n = n_1 + n_2 + \dots + n_k$$

$$S_i = \sum_{j=1}^{n_i} [R(y_{ij})]^2$$

$$\bar{S} = \frac{1}{n} \sum_{i=1}^k S_i$$

$$D^2 = \frac{1}{n-1} \left[\sum_{i=1}^k \sum_{j=1}^{n_i} R(y_{ij})^4 - n(\bar{S})^2 \right] \tag{25.59}$$

If there are no tied ranks, then D^2 and \bar{S} reduce to:

$$D^2 = n(n+1)(2n+1)(8n+11)/180 \tag{25.60}$$

and

$$\bar{S} = (n + 1)(2n + 1) / 6 \quad (25.61)$$

The critical value for the test is $\chi^2_{1-\alpha}(k-1)$, obtained from Table T-2 of the appendix. We reject H_0 at the α level of significance if T exceeds $\chi^2_{1-\alpha}(k-1)$.

Rejection of H_0 indicates that at least two of the populations have unequal variances, but the test does not identify which variances differ. Multiple comparisons of the k populations are made by comparing two variances at a time. We conclude that the variances σ_i^2 and σ_j^2 are different if:

$$\left| \frac{S_i}{n_i} - \frac{S_j}{n_j} \right| > t_{1-\alpha/2}(n-k) \sqrt{D^2 \frac{n-1-t}{n-k} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \quad (25.62)$$

where $t_{1-\alpha/2}(n-k)$ is obtained from Table T-3 of the appendix.

Example 25.10(continued). Comparison of four production lines. We use the k -sample squared ranks test and the data from Example 25.8 to test the hypothesis that the variances of the uranium concentrations of the four production lines are equal. The level of significance of the test is set at $\alpha = 0.10$.

The sample means of the four lines are $\bar{y}_1 = 21.20$, $\bar{y}_2 = 25.86$, $\bar{y}_3 = 18.22$, and $\bar{y}_4 = 26.38$. Table 25.19 lists the absolute differences $|y_{ij} - \bar{y}_i|$, along with their ranks relative to the entire dataset. The bottom row of the table shows r_i , the sum of the squared ranks for the i^{th} sample.

Table 25.19. Ranks for Example 25.10

| Line 1 | | Line 2 | | Line 3 | | Line 4 | |
|------------------------|---------|------------------------|---------|------------------------|---------|------------------------|---------|
| $ y_{1j} - \bar{y}_1 $ | Rank | $ y_{2j} - \bar{y}_2 $ | Rank | $ y_{3j} - \bar{y}_3 $ | Rank | $ y_{4j} - \bar{y}_4 $ | Rank |
| 8.20 | 28.5 | 8.12 | 27 | 1.22 | 4 | 6.62 | 23 |
| 1.80 | 6.5 | 9.88 | 33 | 7.78 | 26 | 3.62 | 12 |
| 1.80 | 6.5 | 5.88 | 21 | 7.22 | 25 | 3.38 | 11 |
| 7.20 | 24 | 9.12 | 31 | 5.22 | 17 | 2.38 | 9 |
| 3.80 | 14 | 8.88 | 30 | 5.78 | 20 | 5.62 | 18.5 |
| 0.80 | 1 | 4.12 | 15 | 3.78 | 13 | 6.38 | 22 |
| 13.80 | 35 | 1.12 | 2 | 10.22 | 34 | 9.38 | 32 |
| 8.20 | 28.5 | 2.12 | 8 | 1.78 | 5 | 5.62 | 18.5 |
| 2.80 | 10 | | | 4.78 | 16 | | |
| 1.20 | 3 | | | | | | |
| $r_1 =$ | 3816.00 | $r_2 =$ | 4413.00 | $r_3 =$ | 3612.00 | $r_4 =$ | 3067.50 |

The calculations are:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} [R(y_{ij})]^4 = 28.5^4 + \dots + 18.5^4 = 11265387$$

$$\bar{S} = \frac{1}{35} (3816.0^2 + 4413.0^2 + 3612.0^2 + 3067.5^2) = 425.96$$

$$D^2 = \frac{1}{34} [11265387 - 35(425.96)^2] = 144559$$

$$\sum_{i=1}^k \frac{S_i^2}{n_i} = \frac{3816.0^2}{10} + \frac{4413.0^2}{8} + \frac{3612.0^2}{9} + \frac{3067.5^2}{8} = 6516317$$

$$T = \frac{6516317 - (35)(425.96)^2}{144559} = 1.15$$

Because $1.15 < \chi^2_{0.90}(3) = 6.25$, H_0 is not rejected. There is no statistical evidence that the uranium concentration variances for the four production lines are different.

It is of interest to compare the outcome of the squared ranks test to its parametric counterpart when normality is assumed. As stated in Section 14.7, however, there are more than one such parametric tests. In this section we use Bartlett's test for homogeneity of variances, given in Section 14.7. From

Equation (14.11), the test statistic is distributed approximately as χ^2 with $\nu = 3$ degrees of freedom and the critical value is $\chi_{0.90}^2 = 6.25$. The calculated statistic is $b = 0.28$. Since $0.28 < 6.25$, we have no statistical evidence to conclude that the variances are different. Thus, the parametric test agrees with the nonparametric k-sample squared ranks test.

25.15 Spearman's test of independence for two populations

The concept of correlation is introduced in Chapter 19 as a measure of the linear relation between two random variables X and Y . However, this measure may be significantly affected by one or more outlying observation (Chapter 26) and thus provide a distorted view of the relation between X and Y . Because a nonparametric correlation can be calculated using the ranks of the observations, the nonparametric correlation will not be unduly affected by such outlying observations.

Assume that the data consist of n pairs of observations (x_i, y_i) , $i = 1, \dots, n$, from two populations, X and Y . The parametric coefficient of correlation for the sample is calculated using Equation (19.7), which is repeated here for convenience as Equation (25.63).

$$R_{XY} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \quad (25.63)$$

where S_{XY} , S_{XX} , and S_{YY} are the working sum of squares formulas:

$$S_{xy} = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i \quad (25.64)$$

$$S_{xx} = \sum x_i^2 - \frac{1}{n} \sum x_i^2 \quad (25.65)$$

$$S_{yy} = \sum y_i^2 - \frac{1}{n} \sum y_i^2 \quad (25.66)$$

where the summations are made from $i = 1$ to $i = n$.

Spearman's rank correlation coefficient

The most commonly used nonparametric measure of correlation is **Spearman's rank correlation coefficient**, which is also referred to as **Spearman's ρ** or **Spearman's**

Spearman's rho

rho It is calculated by essentially replacing the observations by their ranks. Specifically, let $R(x_i)$ denote the rank of x_i compared with the other observations on X , and let $R(y_i)$

denote the rank of y_i compared with the other observations on Y . In case of ties, assign the average of the ranks that would otherwise be assigned if there were no ties.

Spearman's correlation coefficient, denoted by r_s , is defined as:

$$r_s = \frac{\sum_{i=1}^n [R(x_i)R(y_i)]^2 - \frac{n}{4}(n+1)^2}{\sqrt{\sum_{i=1}^n [R(x_i)]^2 - \frac{n}{4}(n+1)^2} \sqrt{\sum_{i=1}^n [R(y_i)]^2 - \frac{n}{4}(n+1)^2}} \quad (25.67)$$

If the data contain no ties, r_s is simplified to:

$$r_s = 1 - \frac{6 \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n(n^2 - 1)} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad (25.68)$$

where $D_i = R(X_i) - R(Y_i)$ is the difference between the ranks of X and Y of the i^{th} pair.

It can be shown (Conover (1980), p. 253) that Equation (25.68) yields the same result as Equation (25.63) when the data are replaced by their ranks. Hence we can use Excel's =CORREL function to calculate r_s if there are no tied data.

Spearman's rho is used to test whether X and Y are related. The null hypothesis is:

$$H_0: X \text{ and } Y \text{ are independent.} \quad (25.69)$$

The alternative hypothesis can be either two-sided or one-sided, depending on the relationship between X and Y . The possibilities are:

$$H_1: \begin{array}{l} \text{There is either a tendency for larger values of } X \\ \text{to be paired with larger values of } Y \text{ or for} \\ \text{larger values of } X \text{ to be paired with smaller} \\ \text{values of } Y. \end{array} \quad (25.70)$$

$$H_1: \text{There is a tendency for larger values of } X \text{ to be} \quad (25.71) \\ \text{paired with larger values of } Y.$$

$$H_1: \text{There is a tendency for larger values of } x \text{ to be} \quad (25.72) \\ \text{paired with smaller values of } Y.$$

If the alternative is given by Equation (25.70), H_0 is rejected if $r_S > w_{1-\alpha/2}$ or if $r_S < w_{\alpha/2}$.

If the alternative is given by Equation (25.71), H_0 is rejected if $r_S > w_{1-\alpha}$.

If the alternative is given by Equation (25.72), H_0 is rejected if $r_S < w_\alpha$.

The quantiles $w_{1-\alpha}$ are given in Table T-19, “Quantiles, $w_{1-\alpha}$, of the Spearman statistic” of the appendix for $n \leq 30$ and for $\alpha = 0.10, 0.05, 0.025, 0.01, 0.005,$ and 0.001 .

When $n > 30$, the $w_{1-\alpha}$ quantile of r_S is calculated as:

$$w_{1-\alpha}(n) = \frac{z_{1-\alpha}}{\sqrt{n-1}} \quad (25.73)$$

where $z_{1-\alpha}$ is given in Table T-1 of the appendix.

For example, when $n = 50$ and $\alpha = 0.05$, the critical value for testing H_0 when the alternative is a positive relationship (Equation (25.71)) is

$$w_{0.95}(50) = 1.645 / \sqrt{49} = 0.235.$$

Example 25.11. Solid waste barrel standards (Bowen and Bennett, p.518). Two uranium-bearing solid waste barrel standards are counted on the same day using the same NDA counting instrument on 12 different days. If there is no shift over time in the bias of the counter, the successive measurements will be independent. However, if the bias of the counter changes over time, the paired measurements on the two barrels will be positively correlated. Test whether there is a bias in the counter at an $\alpha = 0.05$ level of significance.

Table 25.20 shows the 12 measurements (denoted by x_i and y_i), their ranks within each barrel and the rank differences.

Table 25.20. Data and ranks for Spearman’s correlation

| Observation,
<i>i</i> | Raw data | | Ranks | | $D_i =$ | |
|--------------------------|----------|-------|----------|----------|-------------------|-----------|
| | x_i | y_i | $R(x_i)$ | $R(y_i)$ | $R(x_i) - R(y_i)$ | $(D_i)^2$ |
| 1 | 12.92 | 9.76 | 12 | 1 | 11 | 121 |
| 2 | 12.50 | 10.12 | 8 | 8 | 0 | 0 |
| 3 | 12.61 | 10.57 | 10 | 11 | -1 | 1 |
| 4 | 12.48 | 10.80 | 7 | 12 | -5 | 25 |
| 5 | 12.44 | 9.94 | 6 | 3 | 3 | 9 |
| 6 | 12.62 | 10.17 | 11 | 9 | 2 | 4 |
| 7 | 12.35 | 10.08 | 5 | 6 | -1 | 1 |
| 8 | 12.54 | 9.97 | 9 | 5 | 4 | 16 |
| 9 | 12.21 | 10.11 | 2 | 7 | -5 | 25 |
| 10 | 12.22 | 10.22 | 3 | 10 | -7 | 49 |
| 11 | 12.17 | 9.79 | 1 | 2 | -1 | 1 |
| 12 | 12.34 | 9.96 | 4 | 4 | 0 | 0 |
| | | | | | Sum | 252 |

The null hypothesis is given by Equation (25.69) and the alternative by Equation (25.71), Using the intermediate calculations shown in Table 25.20, r_s is calculated using Equation (25.68) as:

$$r_s = 1 - \frac{(6)(252)}{12(12^2 - 1)} = 0.12$$

The critical value for $n = 12$ is found in Table T-19 of the appendix as $w_{0.95}(12) = 0.4965$. Because $0.12 < 0.5804$, we have no statistical evidence of correlation between the two standards, and conclude that there is no shift in the bias of the counter.

In parallel with the nonparametric test of the null hypothesis in Example 25.11, we can also calculate the parametric correlation coefficient R_{XY} . Using Excel function =CORREL, or calculating R_{XY} directly from Equation (19.7), we obtain $r_{XY} = 0.019$. Using Equation (19.9), we can also test the null hypothesis that $\rho = 0$ using the test statistic $t = 0.060$. From Table T-3 of the appendix we have $t_{0.975}(10) = 2.23$. Because $-2.23 < 0.060 < 2.23$, we do not reject H_0 , reaching the same conclusion as Spearman's test.

Kendall's rank correlation coefficient Another nonparametric measure of correlation is **Kendall's rank correlation coefficient**, which is also referred to as **Kendall's τ** or **Kendall's tau**. The interested reader may find details on Kendall's procedure in Conover (1980), p.256, or Bowen and Bennett (1988), p. 521.

Kendall's tau

26

Outliers

26.1 What to look for in Chapter 26

Sometimes it happens that some observations appear to be inconsistent with other observations in the same sample. Such observations may be extreme manifestations of randomness or they may be the result of mistakes or errors or another factor of practical importance. In this chapter, we present procedures to determine whether such apparently erroneous observations should be identified as outliers and removed from the sample when it is analyzed. In this chapter we will:

- *discuss some possible sources of data inconsistency, §26.2*
- *discuss the consequences of a wrong decision to exclude, or fail to exclude, an observation, §26.6*

We present several statistical procedures for identifying outliers:

- *box plot rule, §26.3*
- *Dixon's test, §26.4*
- *Grubbs' tests, §26.5*

26.2 What is an outlier?

There are numerous definition of an outlying observation, or outlier. Barnett and Lewis (1994) define an outlier as an observation that “deviates markedly from other members of the sample in which it occurs.” Hawkins (1980) defines an outlier as an observation that “deviates so much from other observation as to arouse suspicion that it was generated by a different mechanism.” Others define an outlier as one of the set of values that lies unexpectedly distant from most of the other values in the set. The identification of an outlier can be made subjectively, such as in a visual inspection of the observations, or objectively, by using one method from the plethora of methods that are found in the literature. This chapter presents only selected procedures for testing outliers. More extensive procedures are given in books by Barnett and Lewis (1994) and Hawkins (1980), and a survey paper by Beckman and Cook (1983).

outlier In this book, we define an *outlier* as an apparently erroneous observation that has been identified by some statistical procedure as due to error or some cause rather than randomness in the data.

- The removal of an observation is made only for the data analysis. It does not mean throwing away data, as every observation must be recorded. The choice of an outlier rejection procedure should be made with careful consideration of the data structure and the required degree of conservatism. Furthermore, the procedure should be chosen in the planning stage and not after the data are collected.

An observation can deviate from the remaining observations in a sample either because of random fluctuation or because it does not really belong in the sample. If a deviant observation is a legitimate manifestation of randomness but is removed from the sample, the reduced sample is no longer an unbiased sample of the population from which it was drawn. On the other hand, if an observation that does not belong in the sample is retained, the consequent analysis would be tainted. These considerations are discussed in Section 26.6.

There are many possible reasons for an observation to be erroneously included in a sample. These include:

- calibration error
- instrument resolution error
- hysteresis
- error due to under- or over-sensitivity of an instrument in some data ranges
- an unstable scale or instrument
- computation error
- transcription error
- error due to a changing environment, such as a sharp change in temperature

In addition to error, an observation may be the result of malevolent action, such as

- sabotage
 - theft
 - diversion
 - neglect
 - altered records
- Altered records are not usually manifested by outlying observations. On the contrary, records may be altered to cover up extreme values that otherwise may draw attention. It is much harder, if not next to impossible, to identify data that were made to look “normal.”

Once an observation has been identified as an outlier, it is important to examine it to try to identify the reason why it was included in the sample. In some cases, it may be possible to correct the error and restore the observation to the sample. For example, if a decimal point was incorrectly placed, we can correct the record and use the corrected observation. However, if data repair is not possible, such as when there was a contaminated aliquot or an instrument failure, then corrective action might be taken to avoid future errors.

26.3 Box plot procedure for outlier identification

A procedure for identifying outliers in box plots has already been introduced in Section 3.9. There, an outlier (there may be more than one) is defined as any data value that is either:

- smaller than the lower fence, defined as $LQ - 1.5(IQR)$, or
- larger than the upper fence, defined by $UQ + 1.5(IQR)$

where LQ (lower quartile) and UQ (upper quartile) are the 25th and 75th quantiles of the dataset, respectively and IQR (interquartile range) = $UQ - LQ$.

We illustrate the box plot approach to outlier identification in Example 26.1,

Example 26.1. NIST data for outlier identification. NIST (2003), Section 7.1.6, lists a dataset of $n = 90$ values to demonstrate the use of a box plot in outlier identification. These values are listed in Table 26.1.

Table 26.1. NIST data in ascending order of magnitude

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| 30 | 171 | 184 | 201 | 212 | 250 | 265 | 270 | 272 | 289 | 305 | 306 | 322 | 322 | 336 |
| 346 | 351 | 370 | 390 | 404 | 409 | 411 | 436 | 437 | 439 | 441 | 444 | 448 | 451 | 453 |
| 470 | 480 | 482 | 487 | 494 | 495 | 499 | 503 | 514 | 521 | 522 | 527 | 548 | 550 | 559 |
| 560 | 570 | 572 | 574 | 578 | 585 | 592 | 592 | 607 | 616 | 618 | 621 | 629 | 637 | 638 |
| 640 | 656 | 668 | 707 | 709 | 719 | 737 | 739 | 752 | 758 | 766 | 792 | 792 | 794 | 802 |
| 818 | 830 | 832 | 843 | 858 | 860 | 869 | 919 | 925 | 953 | 991 | 1000 | 1005 | 1068 | 1441 |

We calculate the following statistics for this data:

lower quartile, $LQ = 417.25$

upper quartile, $UQ = 748.75$

interquartile range, $IQR = 748.75 - 417.25 = 331.50$

lower fence = $417.25 - 1.5(331.50) = -80.00$

upper fence = $748.75 + 1.5(331.50) = 1246.00$

- As noted in Section 2.6, there are several ways to define and calculate quantiles. Using different definitions result in slightly different calculations of the statistics given above.

The box plot for the data in Table 26.1 is shown in Figure 26.1. The only value identified as an outlier is 1441 because 1441 is greater than 1246, the upper fence.

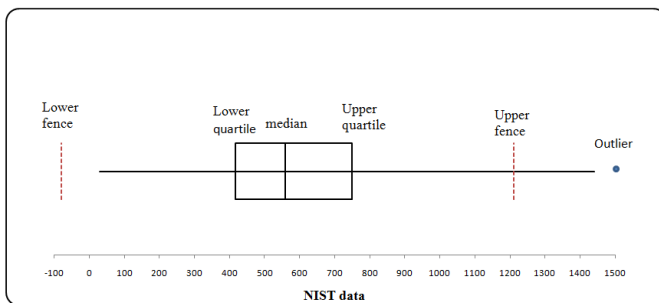


Figure 26.1. An outlier shown in a box plot

26.4 Dixon's procedure for outlier identification

Dixon's test *Dixon's test* for outlier identification (Dixon, 1950, p. 488) is one of the more common procedures developed for this purpose. The procedure assumes that the distribution from which the sample is drawn is normal with unknown parameters and compares the extreme observations in the sample to what might be expected if the distribution were normal.

Dixon's test requires that the n observations y_1, y_2, \dots, y_n in the sample be arranged in ascending order of magnitude. Denote the ordered sample by $\{y_{(1)}, y_{(2)}, \dots, y_{(n)}\}$, with $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$, and where $y_{(i)}$ is the i^{th} order statistic in the sample. In particular, $y_{(1)}$ and $y_{(n)}$, the smallest and largest observations in the sample, respectively, are possible outliers. Dixon's statistic for testing whether either the smallest or the largest value (or both) is an outlier is given in Table 26.2 for sample sizes between 3 and 25.

Table 26.2. Dixon's statistics for identifying outliers

| n | Smallest observation | Largest observation | Designation |
|-------|---|---|-------------|
| 3-7 | $(y_{(2)} - y_{(1)}) / (y_{(n)} - y_{(1)})$ | $(y_{(n)} - y_{(n-1)}) / (y_{(n)} - y_{(1)})$ | r_{10} |
| 8-10 | $(y_{(2)} - y_{(1)}) / (y_{(n-1)} - y_{(1)})$ | $(y_{(n)} - y_{(n-1)}) / (y_{(n)} - y_{(2)})$ | r_{11} |
| 11-13 | $(y_{(3)} - y_{(1)}) / (y_{(n-1)} - y_{(1)})$ | $(y_{(n)} - y_{(n-2)}) / (y_{(n)} - y_{(2)})$ | r_{21} |
| 14-25 | $(y_{(3)} - y_{(1)}) / (y_{(n-2)} - y_{(1)})$ | $(y_{(n)} - y_{(n-2)}) / (y_{(n)} - y_{(3)})$ | r_{22} |

An extreme observation that is identified by Dixon's test as an outlier is erroneously identified if all the sample values in fact came from a normal distribution. Let α be the probability of erroneously labeling an observation as an outlier.

- If the largest observation is suspect, calculate the statistic under the "Largest observation" column for n . If its value is larger than $P_{1-\alpha}$ in Table T-23 of the appendix, then $Y_{(n)}$ is declared an outlier.
 - If the smallest observation is suspect, calculate the statistic under the "Smallest observation" column for n . If its value is larger than $P_{1-\alpha}$ in Table T-23, then $y_{(1)}$ is declared an outlier.
- Note that Dixon's test does not allow both the smallest and the largest observations to be declared outliers.
- An extension of Dixon's table for r_{22} and $26 \leq n \leq 30$ may be found in Bohrer (2008).

Although Dixon's test can identify only one outlier in a single dataset, it can be applied a second time after identifying an outlier in a dataset by eliminating the outlier and then applying the test to the reduced dataset with n replaced by $n - 1$. If an outlier is identified in the reduced dataset, the test can be repeatedly applied until no outliers are identified. However, if Dixon's test is applied more than once to a single dataset, the overall significance of the test is larger than the specified α for the individual test. Tietjen and Moore (1972) address the problem of repeated application of a test procedure to a single dataset.

Example 26.2. Cladding thickness (Bowen and Bennett, 1988, p. 556). Ten ordered observations of cladding thickness (in inches) on fuel rods are 0.0284, 0.0285, 0.0285, 0.0285, 0.0286, 0.0286, 0.0286, 0.0289, 0.0292, and 0.0298. We wish to test the suspicious observation $y_{(10)} = 0.0298$, the largest value in the dataset. For $n = 10$, we compute r_{11} from Table 26.2:

$$r_{11} = \frac{0.0298 - 0.0292}{0.0298 - 0.0285} = 0.4615$$

From Table T-23 of the appendix, the probability that a value of r_{11} larger than 0.4615 would occur by chance is about 0.06, so this observation would be rejected and declared an outlier at the $\alpha = 0.10$ level of significance but not at the $\alpha = 0.05$ level.

26.5 Grubbs' tests for outliers

Grubbs' tests Dixon's test can identify only one outlier in a dataset. Grubbs (1969) developed an assortment of statistics, known as **Grubbs' tests**, that test the largest, the smallest, the two largest, the two smallest, and the largest and smallest observations together. We present these procedures in this section, largely taken from Bowen and Bennett (1988, Section 10.4.2). Grubbs also developed other procedures for far less common cases. These procedures may be found in Grubbs' original paper of 1969 and in Bowen and Bennett (1988, Section 10.4.2)

Grubbs' tests assume that the data come from a normal distribution with unknown parameters. Tests for outliers from a lognormal distribution use the logarithms of the data values. In other non-normal cases, the Grubbs' tests can be used if a transformation of the data to approximate normality can be found.

Statistics used by one or more of Grubbs' procedures are given in Table 26.3.

Table 26.3. Statistics used in Grubbs' procedures

| | | |
|--|---|--------|
| Sample mean | $\bar{y} = \sum_{i=1}^n y_{(i)} / n$ | (26.1) |
| Sample variance | $s^2 = \frac{\sum_{i=1}^n y_{(i)}^2 - \left(\sum_{i=1}^n y_{(i)}\right)^2 / n}{n - 1}$ | (26.2) |
| Sample standard deviation | $s = \sqrt{s^2}$ | (26.3) |
| Sample sum of squares | $SS = (n - 1)s^2$ | (26.4) |
| Sum of squares, omitting two smallest observations | $SS_{1,2} = \frac{\sum_{i=3}^n y_{(i)}^2 - \left(\sum_{i=3}^n y_{(i)}\right)^2}{n - 2}$ | (26.5) |
| Sum of squares, omitting two largest observations | $SS_{n-1,n} = \frac{\sum_{i=1}^{n-2} y_{(i)}^2 - \left(\sum_{i=1}^{n-2} y_{(i)}\right)^2}{n - 2}$ | (26.6) |
| Sample range | $w = y_{(n)} - y_{(1)}$ | (26.7) |

Outlier scenarios that are handled by Grubbs' tests are given in Table 26.4.

Table 26.4. Outlier scenarios and test Statistics (Bowen and Bennet, 1988, p .558)

| Task | Statistic | Comment |
|---|---------------------------------------|---|
| Testing the largest value as an outlier | $T_n = \frac{y_{(n)} - \bar{y}}{s_v}$ | Critical values are provided in Table T-20 of the appendix for $\alpha = 0.001, 0.005, 0.01, 0.025, 0.05,$ and $0.10,$ for $n = 3$ to $n = 140.$ If there is no reason for testing either a largest or a smallest value as an outlier, the actual significance levels are $0.002, 0.01, 0.02, 0.05, 0.10,$ and $0.20.$ |
| Testing the smallest value as an outlier | $T_1 = \frac{\bar{y} - y_{(1)}}{s_v}$ | |
| Testing jointly the two extremes as outliers | w/s | Critical values for the ratio of the sample range and standard deviation are given in Table T-21 of the appendix for $\alpha = 0.005, 0.01,$ and 0.05 and selected values of $n.$ A value of w/s that exceeds a critical value is statistically significant at that level. If significance is found, the smallest and largest values should be examined to determine the reason for significance. |
| Testing jointly the two smallest observations as outliers | $SS_{1,2} / SS$ | Small values of the statistic indicate significance. Critical values are found in Table T-22 of the appendix for $\alpha = 0.001, 0.005, 0.01, 0.025, 0.05,$ and $0.10,$ for $n = 3$ to $n = 140.$ |
| Testing jointly the two largest observations as outliers | $SS_{n-1, n} / SS$ | |

Example 26.3. Testing the largest observation in a sample. The statistics T_1 and T_n from Table 26.4 are frequently recommended for testing extreme observations, particularly if n is relatively large. For the data in Example 26.1, the sample mean is $\bar{y} = 0.02876$ and the sample standard deviation is $s = 0.000435$, so that $T_{10} = (0.0298 - 0.02876)/0.000435 = 2.391$.

There was no prior reason to test the largest as opposed to the extreme deviation in either direction, and, accordingly, we need to run a two-sided test. Table T-20 gives critical points for a one-sided test and therefore we have to look up the critical point under $\alpha/2 = 0.05/2 = 0.025$. For $n = 10$, the critical value for

T_{10} is 2.290. Because $2.391 > 2.290$ we identify $y_{(10)} = 0.0298$ as an outlier. Note that Dixon's test, applied to the same sample, does not identify $y_{(10)}$ as an outlier. This suggests that if the data are normal or nearly so, Grubbs' test is more sensitive to the presence of outliers.

26.6 Considerations in outlier rejection

When a statistical test identifies an outlier, there may be compelling reasons to remove the observation from the sample. There may also be equally compelling arguments to keep the outlier in the sample. In this section, we consider the consequence of a decision to reject or retain an outlier. Some of the arguments presented here are intuitively obvious and some may merit further consideration.

An observation that is far removed from the remaining observations in the sample could be the result of scenarios such as:

- a contaminated datum, such as a chemical contamination of an aliquot
- an instrument that failed during an experiment, such as during a power interruption
- a survey of employees' salaries that included the company president's salary
- a large credit card expense due to fraud
- an extremely high electric bill due to incorrect reading of the meter
- a drastically low inventory due to theft or diversion
- an extremely high repair charge due to swindling

In each of these cases, an analysis of the data that includes the outlier may lead to the wrong conclusion about the population of interest. The researcher should investigate each apparently erroneous observation (whether identified as an outlier or not) for cause. If a source of data contamination is identified, the researcher must determine whether to include or exclude the questionable observation. In either case, the results of the investigation should be reported, including the rationale to include or exclude the observation of concern.

If a source of data contamination cannot be identified, the decision to include or exclude an apparently erroneous observation is much harder. If an outlier is excluded from the analysis (a "sin of omission"), this would result in a smaller standard deviation, or some other measure of variability, than the data show. This may suggest that we have tighter control of the experiment, or a shorter confidence interval than is appropriate, and lead to an incorrect conclusion about the population from which the sample is drawn.

On the other hand, if we wrongly leave the questionable observation in the sample (a "sin of commission"), this may lead to biased results in the reported statistics and the associated inference. The increased standard deviation would

lead to a longer confidence interval that covers a wider range of possibilities and is less likely to reject a null hypothesis.

In taking account of the potential cost of an incorrect decision, many experienced researchers usually prefer to include questionable, but plausible, observations in the analysis. The writers of this book do not make such a recommendation. However, we do suggest that researchers run the analysis twice: once with and once without the outliers. If the conclusions differ, then the more conservative (in the public's interest) conclusion should be adopted. The final report, however, must indicate that two analyses were performed and explain why the report's conclusion was adopted.

27

Simulation

27.1 What to look for in Chapter 27

This chapter discusses:

- *the role of simulation, §27.2*
- *generation of random numbers, §27.3*
- *estimation of definite integrals, §27.4*
- *generation of normal variates, §27.5*
- *generation of arbitrary variates, §27.6*
- *confirmation of the central limit theorem, §27.7*

27.2 Introduction to simulation

simulation **Simulation** is the process of modeling and analyzing random events using random variables. Those random variables may be obtained from a variety of sources, such as a table of random

numbers, a flip of a coin, or radioactive decay. Sources of random numbers include hand-held calculators, laptops, home computers, and mainframes.

Monte Carlo Simulation is often performed using a **Monte Carlo** procedure, which is based on the observed fraction of random events satisfying a property of interest. The term *Monte Carlo* was coined by S. Ulam, a mathematician who worked on the Manhattan project in Los Alamos. It is named after the famous Monte Carlo gambling casino in Monaco, on the shore of the Mediterranean. There, using paper and pencil, gamblers observed and recorded long series of cards, dice, and roulette wheel behavior to determine a possible trend that would give them an edge in their gambling.

history We may conduct a simple Monte Carlo study by repeating an experiment many times. Each repetition, called a **history**, must be conducted in an identical manner. For example, we could roll a pair of fair dice 1000 times and count the number of times that the sum of the dots is 7 or 11. Using this count, we can obtain an empirical estimate of the probability $Pr\{7 \text{ or } 11\}$ without resorting to mathematical calculations. For example, if we rolled 7 or 11 in 211 out of 1000 histories, then our empirical estimate of $Pr\{7 \text{ or } 11\}$ is $211/1000 = 0.211$. Our intuition also tells us that if we increase the number of histories, our estimate would be more precise.

Simulation procedures have many applications in statistics, quality control, and many other sciences. For example, we may use simulation methods to:

- estimate the probability of an event for which the distribution is unknown or unspecified
- estimate quantiles of a distribution
- examine the validity of assumptions about a specific distribution
- estimate parameters when analytical methods are not available
- examine the claim of the probability of Type I and Type II errors
- compare sample estimates with the known parameters of a distribution
- select random combinations of treatment levels in multi-factor experiments when we cannot experiment with all combinations of all levels of all factors
- verify that a procedure for selecting employees for drug testing is random .
- validate (rather than prove) mathematical relations, such as the central limit theorem, without resorting to advanced calculus methods
- determine which Blackjack strategy to use on our next visit to Las Vegas

System simulation may be used to create a model that duplicates the process (or system) behavior. By manipulating the model, we can gain a better understanding of the process, examine “what if” departures from a specified model, and determine the effect of such departures. Many organizations, including the NRC, have been developing such models to describe and analyze safety, production, measurement, and quality control processes.

Most simulations require a computer and one or more software routines to produce the appropriate random variables, record the outcomes, and perform the necessary calculations. We can imagine that even the simple case of estimating $Pr\{7 \text{ or } 11\}$ by rolling a pair of fair dice 1000 times may take hours to roll, count, record, and calculate. A home computer may simulate the process and provide the results in seconds.

Some simple simulations may be performed using Excel. Simulations that are more complicated may require a dedicated software package. In this book, we use Excel’s functions exclusively.

It is commonly known that no algorithm produces a truly random number. However, some algorithms can generate numbers that are “close enough” to randomness for use in various investigations. Excel and many other software packages can generate random samples from the uniform distribution $U(0,1)$ that may be sufficient for many applications. Some software packages (not including Excel) allow a user to input of a “seed number” to allow a user to duplicate the sequence of variates, if necessary for validation purposes.

The requirements for Monte Carlo simulations include (Sawilowsky, 2003):

- a (pseudo) random number generator that has been tested extensively and verified using statistical tests for randomness
- a random number generator that has a long sequence before it repeats itself
- a sufficiently large number of histories to assure reliable results
- a planned and proper sampling technique to assure that the study simulates the distribution or the phenomena of interest

27.3 Computer generator of random numbers

The generation of random numbers can be thought of as the statistical equivalent of throwing darts at a target.

variate Most modern computers have one or more routines to generate random numbers, called *variates*. The most basic computer routine generates variates from the uniform $U(0, 1)$ distribution, which is described in Section 7.2. Excel has the following routines and options for such generation:

=RAND() returns a variate between 0 and 1. Note the empty parentheses! Excel's =RAND function is volatile! Thus, after we call this function from cell A1, just about any operation in the spreadsheet will cause the content of cell A1 to change. The only way to keep a random number from changing is to use the Paste Special command and copy a cell (or cells) to itself as a "value." This procedure is illustrated below where we wish to generate and keep ("freeze") 20 such variates in cells A1..A20.

- Place the cursor in cell A1 and enter =RAND().
- Copy cell A1 to A1..A20.
- Highlight cells A1..A20.
- Place the cursor in cell A1.
- Click on **copy**
- Click on **Paste**, select **Paste Special**, select **Values**, and **OK**. On some versions of Excel, we call the Paste Special command from the Edit panel.

=RAND()*($b - a$) + a returns a variate between a and b , where a and b are any two real numbers and $a < b$. For example, to generate a random number between number between 75 and 79, call =RAND()*(79-75)+75.

=RANDBETWEEN (I_1, I_2) returns a variate integer I between I_1 and I_2 , inclusively. This function is especially helpful in generating random from discrete distribution of consecutive integers. For example, if we wish to randomly select a record from set of records that are numbered 103 to 284, we call =RANDBETWEEN(103, 284) from cell A1. If we need 10 records from the same set. we copy cell A1 to A2..A10.

Note that the operation just described may yield duplicates. To assure that we have 10 distinct records, we need to generate some extra records and replace each duplicate by the next available record from the list of extras. Such list is useful when we call employees for drug testing or potential jurors for jury duty. If the randomly selected individual is unavailable, we pick the next available individual from the list of extras.

27.4 Evaluating definite integrals

In this section we illustrate some of the uses of the $U(0, 1)$ variates in simulation to address specific problems. The first illustration shows how simulation can approximate the definite integral:

$$\int_a^b f(y) dy \quad (27.1)$$

where $f(y) \geq 0$ is a known non-negative function. Then I is the area under the curve $f(y)$ for $a \leq y \leq b$. (See Figure 27.1.)

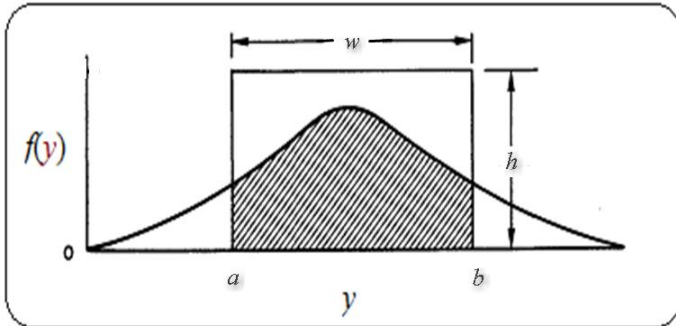


Figure 27.1. Estimation of a definite integral by simulation

To estimate Equation (27.1) by simulation, start by constructing a rectangle that frames the function in the region of the integration. The base of the rectangle is a horizontal line from a to b , the limits of integration, with width $w = b - a$. Choose the height h of the rectangle so that the function $f(y)$ is inside the rectangle over the interval $a \leq y \leq b$. The area of the rectangle is then

$A = wh = (b - a)h$. This is illustrated in Figure 27.1.

Next we choose random points within the rectangle. Let M be the number of random points chosen, i.e., the number of histories, and let m be the number of times the random point falls in the shaded area in Figure 27.1. An unbiased estimate of I is then:

$$\hat{I} = (m/M) A \quad (27.2)$$

The process just described is detailed in the following steps using Excel:

Step 1. Use a $U(0, 1)$ generator, such as `RAND()`, to obtain u_{11} , a $U(0,1)$ variate.

- Step 2.** Calculate $v_{11} = u_{11}(b - a) + a$, a uniform random variate from the interval (a, b) . This is the abscissa of the first random point in the rectangle.
- Step 3.** Use an independent $U(0, 1)$ generator to generate u_{12} , a second $U(0,1)$ variate.
- Step 4.** Calculate $v_{12} = u_{12}h$, a uniform variate from the interval $(0, h)$. This is the ordinate of the first random point in the rectangle, which is (v_{11}, v_{12}) .
- Step 5.** Calculate $f(v_{11})$, the height of $f(y)$ at $y = v_{11}$.
- Step 6.** Let $c_1 = 1$ if $v_{12} < f(v_{11})$, and 0 otherwise.
- Step 7.** Repeat Steps 1 through 6 M times, where M is the number of histories planned for the study, and determine $c_1, c_2, \dots, c_i, \dots, c_M$.
- Step 8.** Let $m = \sum c_i$ = the number of histories where the random point (v_{11}, v_{12}) falls in the shaded area of Figure 27.1.
- Step 9.** Calculate the hit ratio m/M and multiply that ratio by the area A of the rectangle to obtain an estimate of \hat{I} , per Equation (27.2).

In many applications, $f(y)$ in Equation (27.1) is the density of a random variable Y , so that I is the probability that $a < Y < b$. This is illustrated by the following example.

Example 27.1. Evaluation of a probability by simulation. Following Bowen and Bennett, p. 580, we suppose Y is a standard $N(0,1)$ variable and we wish to estimate the $Pr\{-1.3 < Y < 1.6\}$. This probability can be written as

$$Pr\{-1.3 < Y < 1.6\} = \int_{-1.3}^{1.6} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \quad (27.3)$$

This probability is the shaded area in Figure 27.1, with $a = -1.3$ and $b = 1.6$. For Example 27.1, the width of the rectangle is $w = 1.6 - (-1.3) = 2.9$. The maximum value of $f(y)$ over the interval $-1.3 \leq y \leq 1.6$ occurs at $y = 0$, where $f(y) = 1/\sqrt{2\pi} = 0.3989$. Thus, the height of the framing rectangle must exceed 0.3989. For this example, we chose $h = 0.5$, and the area of the rectangle is $A = 2.9(0.5) = 1.45$. For Steps 2 and 4, we have

$$v_{11} = u_{11}[1.6 - (-1.3)] + (-1.3) = u_1(2.9) - 1.3 \text{ and } v_{12} = (u_{12})(0.5).$$

Bowen and Bennett ((1988), p. 581) generate $M = 1000$ histories and show the first five histories. We repeated the experiment, also with 1000 histories and our first five histories are shown in table 27.1. Of course, our histories differ from Bowen and Bennett's because our random numbers are different.

Table 27.1. First five histories in simulation

| u_1
(Step 1) | v_1
(Step 2) | u_2
(Step 3) | v_1
(Step 4) | $f(v_1)$
(Step 5) | $c: v_2 < f(v_1)$
(Step 6) |
|-------------------|-------------------|-------------------|-------------------|----------------------|-------------------------------|
| 0.585171 | 0.396996 | 0.546777 | 0.273389 | 0.368711 | 1 |
| 0.415463 | -0.095159 | 0.637761 | 0.318880 | 0.397140 | 1 |
| 0.874140 | 1.235006 | 0.628516 | 0.314258 | 0.186084 | 0 |
| 0.313323 | -0.391364 | 0.888004 | 0.444002 | 0.369531 | 0 |
| 0.138010 | -0.899771 | 0.173597 | 0.086779 | 0.266140 | 1 |

In our simulation, Excel counted 578 histories where $v_{i2} < f(v_{i1})$ so that $c_i = 1$. The estimate of $Pr\{-1.3 < Y < 1.6\}$ is therefore $(578)(1.45)/1000 = 0.8381$.

Note that Bowen and Bennett's simulation estimated this probability as $(601)(1.45)/1000 = 0.8715$. The average of the two estimates, which is equivalent to a simulation with $M = 2000$ histories, is 0.8548.

From Table T-1 of the appendix, the actual value of the probability is:

$$\Phi(1.6) - \Phi(-1.3) = 0.9452 - 0.0968 = 0.8484.$$

We see that the errors in the two simulation estimates are about 0.01 and 0.02, while the error in the average is about 0.005. Although increasing M does lead to a smaller error for this example, this is not always the case. However, increasing M does decrease the variance of the estimate. Because the simulation estimate is unbiased, this would usually decrease the error.

There is another simulation method for estimating the definite integral in Equation (17.1) that may be more accurate in some cases. It is based on the approximation given by

$$\int_a^b f(y)dy \cong \frac{b-a}{M} \sum_{i=1}^M f(y_i) \quad (27.4)$$

The method is to replace y_i by a uniform random variate v_i from the interval (a, b) as in Step 2 above and then sum the M values of $f(v_i)$.

To apply this second method to Example 27.1, we generated $M = 1000$ histories, each yielding a $U(-1.3, 1.6)$ variate, using Excel's function $=\text{RAND}()*(1.6 + 1.3) - 1.3$. We summed the 1000 $f(v_i)$ values and, multiplied the sum by $(1.6 + 1.3)/1000$ to obtain an estimate of 0.8503. Bowen and Bennett's simulation (p. 582) returned the value of 0.8423.

27.5 Generation of normal variates

Simulation studies often require input of observations from a normal distribution. There are several computer routines for generation of normal variates, two of which are given below. Both procedures are based on a transformation of uniform $U(0, 1)$ variates.

Procedure 1. This procedure utilizes a version of the central limit theorem that states that the sum of n independent variates $\{Y_i\}$ having a distribution with mean μ and standard deviation σ is approximately distributed as a normal variate with a mean of $n\mu$ and standard deviation of $\sigma\sqrt{n}$. (See Table 7.2 of Section 7.8.) Mathematically stated (when n is large):

$$\sum_{i=1}^n Y_i \approx N(n\mu, n\sigma^2) \quad (27.5)$$

From Equation (7.9), the mean and the variance of a standard uniform variate $U(0, 1)$ are:

$$\mu = 1/2 \quad (27.6)$$

$$\sigma^2 = 1/12 \quad (27.7)$$

Let $\{Y_i\}$ be a set of n independent standard uniform variates. The convergence to a normal distribution in Equation (27.5) occurs very quickly so that for $n = 12$, we have:

$$\sum_{i=1}^{12} Y_i \approx N((12)(0.5), 12/12) = N(6, 1) \quad (27.8)$$

Therefore, by subtracting 6 from the sum of 12 $U(0,1)$ variates we can generate an excellent approximation to a standard normal $N(0,1)$ variate. Neglecting the error in the approximation, we have:

$$Z = \sum_{i=1}^{12} Y_i - 6 \sim N(0, 1) \quad (27.9)$$

➤ Procedure 1 uses 12 $U(0, 1)$ variates to obtain a single $N(0, 1)$ variate.

Example 27.2. Generating $N(0,1)$ variates by sum of $U(0,1)$ variates.

Using Excel, we generated the values of $\{0.2884, 0.6764, 0.2043, 0.3431, 0.0193, 0.7441, 0.6460, 0.7603, 0.3034, 0.4701, 0.9174, 0.8487\}$. The sum of the 12 values is 6.2214, and the generated $N(0,1)$ is set at $6.2215 - 6 = 0.2215$.

Box-Muller transformation

Procedure 2. This procedure for generating $N(0, 1)$ involves trigonometric transformations. Called the **Box-Muller transformation** (Box and Muller (1958)), this procedure transforms a pair U_1 and U_2 of independent $U(0,1)$ variates into a pair Z_1 and Z_2 of $N(0, 1)$ variates:

$$Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2) \quad (27.10)$$

$$Z_1 = \sqrt{-2 \ln U_1} \sin(2\pi U_2) \quad (27.11)$$

The variates Z_1 and Z_2 generated using Equations (27.9) and (7.10) are independent $N(0, 1)$ variates.

- The Box-Muller transformation requires n $U(0,1)$ variates to generate n $N(0, 1)$ variates if n is even and $n + 1$ variates if n is odd. This is considerably less than the $12n$ variates required by Procedure 1.

Example 27.3. Generating normal variates using the Box-Muller transformation. Suppose a pair of $U(0, 1)$ variates is $u_1 = 0.7374$ and

$u_2 = 0.4414$. We have: $-2 \ln(u_1) = 0.6092$, and $\sqrt{0.6092} = 0.7805$. Next we calculate:

$$2\pi(u_2) = 2.7734, \cos(2.7734) = -0.9330, \sin(2.7734) = 0.3599$$

The pair of normal $N(0, 1)$ variates are:

$$z_1 = (0.7805)(-0.9330) = -0.7282, \quad z_2 = (0.7805)(0.3559) = 0.2809$$

Example 27.3. Generating normal variates using the Box-Muller transformation. Suppose a pair of $U(0, 1)$ variates is $u_1 = 0.7374$ and

$u_2 = 0.4414$. We have: $-2 \ln(u_1) = 0.6092$, and $\sqrt{0.6092} = 0.7805$. Next we calculate:

$$2\pi(u_2) = 2.7734, \cos(2.7734) = -0.9330, \sin(2.7734) = 0.3599$$

The pair of normal $N(0, 1)$ variates are:

$$z_1 = (0.7805)(-0.9330) = -0.7282, \quad z_2 = (0.7805)(0.3559) = 0.2809$$

27.6 Generation of arbitrary variates There is a general procedure for generating a variate from any distribution for which the inverse function F^{-1} of its *cdf* F is known. Let W have a standard uniform $U(0, 1)$ distribution. Then:

$$Y = F^{-1}(W) \quad (27.12)$$

has a distribution with *cdf* F .

- The *cdf* F can be discrete or continuous.

- If only an approximation to F^{-1} is known, then the *cdf* of Y from Equation (27.12) can be used to approximate F .

Excel has several inverse functions that may be used to generate variates in this way. To use the procedure, we call `RAND()` to return a $U(0,1)$ variate, as illustrated below:

Table 27.2. Excel's inverse function for variate generation

| Distribution | Call to inverse function | Examples |
|--------------------------------|--|--|
| $N(0, 1)$ | <code>=NORMSINV(RAND())</code> | <code>=NORMSINV(0.999) = 3.719</code>
<code>=NORMSINV(0.50) = 0.000</code>
<code>=NORMSINV(0.1234) = -1.158</code> |
| $N(\mu, \sigma^2)$ | <code>=NORMINV(RAND(), μ, σ)</code> | <code>=NORMINV(0.5, 10, 2) = 10.00</code>
<code>=NORMINV(0.1234, 10, 2) = 7.684</code> |
| Chi-square (ν) | <code>=CHINV((RAND(), ν))</code> | <code>=CHINV(0.05, 1) = 3.841</code>
<code>=CHINV(0.1234, 7) = 11.367</code> |
| Student- t (ν) | <code>=TINV((RAND(), ν))</code> | <code>=TINV(0.05, 500) = 1.965</code>
<code>=TINV(0.1234, 5) = 1.851</code> |
| $F(\nu_1, \nu_2)$ | <code>=FINV((RAND(), ν_1, ν_2))</code> | <code>=FINV(0.1234, 8, 1) = 38.905</code>
<code>=FINV(0.1234, 1, 8) = 2.965</code> |
| Log-normal (μ, σ^2) | <code>=LOGINV(RAND(), μ, σ)</code> | <code>=LOGINV(0.1234, 0, 1) = 0.314</code>
<code>=LOGINV(0.1234, 2, 0.5) = 4.141</code> |

The procedure can also be implemented graphically from a plot of the *cdf*. This is illustrated in Figure 27.2 for a hypothetical *cdf* plot. Generate a $U(0,1)$ variate to determine a random point w along the vertical axis. Then draw a horizontal line with height w until it intersects the *cdf* at the point (y, w) . In Figure 27.2, $(y, w) = (0.48, 0.655)$. Then y is a variate from a distribution with *cdf* F .

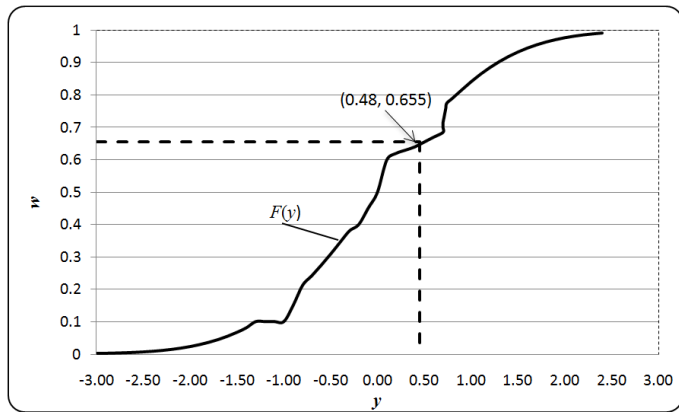


Figure 27.2. Generation of a variate with an arbitrary cdf F

Information on random number generators may be found in Kennedy and Gentle (1980). This book provides algorithms for the generation of variates from many distributions as well as detailed FORTRAN routines to execute these algorithms.

27.7 Confirmation of the central limit theorem

As an example of simulation, we can confirm the central limit theorem by repeating an experiment many times and show that the results are consistent with the theorem. Although not the same as a mathematical proof of the theorem, this is a useful illustration of simulation.

From Section 7.8, the central limit theorem states that if a random sample Y_1, Y_2, \dots, Y_n of size n is drawn from a population with a mean μ and variance σ^2 , then the distribution of the sample mean \bar{Y} is approximately normal:

$$\bar{Y} \approx N(\mu, \sigma^2/n) \quad (27.13)$$

Example 27.4. Confirmation of the central limit theorem by simulation.

As a class exercise in a statistics course, each of k students collects an independent sample of n observations from the same population with a known mean μ and variance σ^2 and calculates the sample mean \bar{y}_i , $i = 1, 2, \dots, k$. (As a separate confirmation exercise, each student also calculates the sample variance S_i^2 .) According to the central limit theorem, $\{\bar{y}_1, \dots, \bar{y}_k\}$ is a set of k observations from a population with distribution approximately given by Equation (27.13). To confirm this, we ask the following questions:

- (1) Is the average $\bar{\bar{y}}$ of $\{\bar{y}_1, \dots, \bar{y}_k\}$ consistent with μ ?
- (2) Is the sample variance s^2 of $\{\bar{y}_1, \dots, \bar{y}_k\}$ consistent with σ^2/n ?

(3) Is the empirical distribution function of the k observations $\{\bar{y}_1, \dots, \bar{y}_k\}$ approximately normal?

Suppose there are $k = 30$ students in the class and each draws $n = 100$ observations from the $U(0,1)$ distribution. Then $\mu = 0.5$ and $\sigma^2 = 1/12 = 0.0833$.

The results of such a simulation are shown in Table 27.3, where the sample mean and variance is given for each of the $k = 30$ samples.

(4) Do the means appear to follow the normal distribution?

We simulate the experiment by drawing $k = 30$ samples (representing 30 students) from the $U(0,1)$ distribution. We selected the $U(0,1)$ distribution because (1) it is easy to generate, (2) we know that the distribution is not normal, and (3) we know the mean ($\mu = 0.5$) and the variance ($\sigma^2 = 1/12 = 0.08333$).

The results of such simulation are shown in Table 27.3, where the sample mean and variance is given for each of the $k = 30$ samples.

Table 27.3. Summary of simulations for Example 27.3

| Sample | Mean, \bar{y} | Variance, S^2 | Sample | Mean, \bar{y} | Variance, S^2 |
|--------|-----------------|-----------------|--------|-----------------|-----------------|
| 1 | 0.4809 | 0.0912 | 16 | 0.5094 | 0.0675 |
| 2 | 0.4861 | 0.0867 | 17 | 0.4637 | 0.0901 |
| 3 | 0.4916 | 0.0835 | 18 | 0.5033 | 0.0835 |
| 4 | 0.5373 | 0.0859 | 19 | 0.5227 | 0.0803 |
| 5 | 0.4840 | 0.0883 | 20 | 0.5207 | 0.0882 |
| 6 | 0.5947 | 0.0719 | 21 | 0.5060 | 0.0821 |
| 7 | 0.5026 | 0.0901 | 22 | 0.4858 | 0.0830 |
| 8 | 0.4969 | 0.0860 | 23 | 0.5511 | 0.0851 |
| 9 | 0.4837 | 0.0842 | 24 | 0.4715 | 0.0852 |
| 10 | 0.4936 | 0.0790 | 25 | 0.5381 | 0.0881 |
| 11 | 0.4990 | 0.0786 | 26 | 0.4931 | 0.0933 |
| 12 | 0.4999 | 0.0771 | 27 | 0.5510 | 0.0763 |
| 13 | 0.4804 | 0.0754 | 28 | 0.5102 | 0.0869 |
| 14 | 0.5071 | 0.0864 | 29 | 0.4700 | 0.0887 |
| 15 | 0.4838 | 0.0719 | 30 | 0.4745 | 0.0769 |

The answers to the questions above are as follows:

(1) The average of $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{30}$ is $\bar{\bar{y}} = 0.5012$. This is in very close agreement with $\mu = 0.5$.

(2) The sample variance of $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{30}$ is $s_y^2 = 0.000654$. This is in very close agreement with $\sigma^2/n = 0.0833/100 = 0.000833$.

- (3) We use the W -test of Section 11.9 to test the normality of the sample means $\{\bar{y}_i\}$. From Equation (11.14), the test statistic is $w = 0.941$. The critical value for $\alpha = 0.05$ is $w_{0.95}(30) = 0.927$. Since $0.941 > 0.927$, we do not reject the null hypothesis of normality and conclude that the statistical evidence is consistent with a normal distribution for the sample means.

Because the answers to all three questions support the central limit theorem, we conclude that this example confirms the central limit theorem.

As a separate confirmation, we check that the sample variances

$s_1^2, s_2^2, \dots, s_{30}^2$ listed in Table 27.3 are consistent with the population variance.

The average of the 30 sample variances is 0.0831, which is in very close agreement with the population variance $\sigma^2 = 0.0833$.

Appendix

Statistical tables

| | |
|---------------------|---|
| Table T-1: | Cumulative standard normal distribution $\Phi(z)$ |
| Table T-2: | Quantiles, $\chi^2_q(v)$, for the chi-square distribution with v degrees of freedom |
| Table T-3: | Quantiles, $t_q(v)$, for Student's t - distribution with v degrees of freedom |
| Table T-4: | Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator |
| Table T-5: | Quantiles, $f_{max, q}(k, v)$, for the F_{max} distribution . |
| Table T-6a: | Coefficients $\{a_{n-i+1}\}$ for the W -test for normality |
| Table T-6b: | Quantiles, $w_q(n)$, for the W -test for normality |
| Table T-7: | Significant ranges, $q_{0.95}(p, N - k)$, for Duncan's multiple-range test |
| Table T-8: | Selected binomial probabilities |
| Table T-9: | Selected Poisson probabilities |
| Table T-10: | Confidence limits for the Poisson parameter λ |
| Table T-11a: | Two-sided tolerance limit factors for a normal distribution |
| Table T-11b: | One-sided tolerance limit factors for a normal distribution |

| | |
|--------------------|--|
| Table T-12: | Two thousand random digits |
| Table T-13: | Quantiles, $d_q(n)$, of the Kolmogorov test statistic |
| Table T-14: | Quantiles, $D'_q(n)$, of the distribution of the D' statistic |
| Table T-15: | Cumulative probabilities $Pr[U \leq u]$ of the total number of runs u in a sample of size n_1+n_2 |
| Table T-16: | Quantiles, W_q , of the Wilcoxon signed ranks test statistic |
| Table T-17: | Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum (WRS) Statistic |
| Table T-18: | Quantiles of the squared ranks test statistic |
| Table T-19: | Quantiles, w_p , of the Spearman statistic, r_s |
| Table T-20: | Critical Values for Grubbs' T -test |
| Table T-21: | Critical Values (one-sided test) for ratio of w/s range to sample standard deviation |
| Table T-22: | Critical Values for $S_{2n-1,n}/S_2$, or $S_{21,2}/S_2$ for simultaneously testing the two largest or two smallest observations |
| Table T-23: | Dixon's criteria for testing for extreme values |

Table T-1: Cumulative standard normal distribution $\Phi(z)$ (see Section 7.6)

| $z \Phi(z)$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|------------------|-----------------------|------------|-------------|------------|-------------|--------------|-------------|-------------|--------------|--------------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |
| Selected | q: | 0.5 | 0.75 | 0.9 | 0.95 | 0.975 | 0.98 | 0.99 | 0.995 | 0.999 |
| quantiles | z_q: | 0.000 | 0.674 | 1.282 | 1.645 | 1.960 | 2.054 | 2.326 | 2.576 | 3.090 |

Table T-2: Quantiles, $\chi^2_q(\nu)$, for the chi-square distribution with ν degrees of freedom (see Section 7.10)

| ν | q | | | | | | | | | |
|-------|--------|--------|--------|--------|--------|-------|-------|-------|-------|-------|
| | 0.005 | 0.010 | 0.025 | 0.050 | 0.100 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 |
| 1 | 0.0000 | 0.0001 | 0.0009 | 0.0039 | 0.0158 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.103 | 0.211 | 4.61 | 5.99 | 7.38 | 9.21 | 10.6 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.25 | 7.81 | 9.35 | 11.3 | 12.8 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.78 | 9.49 | 11.1 | 13.3 | 14.9 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.61 | 9.24 | 11.1 | 12.8 | 15.1 | 16.7 |
| 6 | 0.676 | 0.872 | 1.24 | 1.64 | 2.20 | 10.6 | 12.6 | 14.4 | 16.8 | 18.5 |
| 7 | 0.989 | 1.24 | 1.69 | 2.17 | 2.83 | 12.0 | 14.1 | 16.0 | 18.5 | 20.3 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 13.4 | 15.5 | 17.5 | 20.1 | 22.0 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 14.7 | 16.9 | 19.0 | 21.7 | 23.6 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 16.0 | 18.3 | 20.5 | 23.2 | 25.2 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 17.3 | 19.7 | 21.9 | 24.7 | 26.8 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 18.5 | 21.0 | 23.3 | 26.2 | 28.3 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 19.8 | 22.4 | 24.7 | 27.7 | 29.8 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 21.1 | 23.7 | 26.1 | 29.1 | 31.3 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 22.3 | 25.0 | 27.5 | 30.6 | 32.8 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 23.5 | 26.3 | 28.8 | 32.0 | 34.3 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.1 | 24.8 | 27.6 | 30.2 | 33.4 | 35.7 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.9 | 26.0 | 28.9 | 31.5 | 34.8 | 37.2 |
| 19 | 6.84 | 7.63 | 8.91 | 10.1 | 11.7 | 27.2 | 30.1 | 32.9 | 36.2 | 38.6 |
| 20 | 7.43 | 8.26 | 9.59 | 10.9 | 12.4 | 28.4 | 31.4 | 34.2 | 37.6 | 40.0 |
| 21 | 8.03 | 8.90 | 10.3 | 11.6 | 13.2 | 29.6 | 32.7 | 35.5 | 38.9 | 41.4 |
| 22 | 8.64 | 9.54 | 11.0 | 12.3 | 14.0 | 30.8 | 33.9 | 36.8 | 40.3 | 42.8 |
| 23 | 9.26 | 10.2 | 11.7 | 13.1 | 14.8 | 32.0 | 35.2 | 38.1 | 41.6 | 44.2 |
| 24 | 9.89 | 10.9 | 12.4 | 13.8 | 15.7 | 33.2 | 36.4 | 39.4 | 43.0 | 45.6 |
| 25 | 10.5 | 11.5 | 13.1 | 14.6 | 16.5 | 34.4 | 37.7 | 40.6 | 44.3 | 46.9 |
| 26 | 11.2 | 12.2 | 13.8 | 15.4 | 17.3 | 35.6 | 38.9 | 41.9 | 45.6 | 48.3 |
| 27 | 11.8 | 12.9 | 14.6 | 16.2 | 18.1 | 36.7 | 40.1 | 43.2 | 47.0 | 49.6 |
| 28 | 12.5 | 13.6 | 15.3 | 16.9 | 18.9 | 37.9 | 41.3 | 44.5 | 48.3 | 51.0 |
| 29 | 13.1 | 14.3 | 16.0 | 17.7 | 19.8 | 39.1 | 42.6 | 45.7 | 49.6 | 52.3 |
| 30 | 13.8 | 15.0 | 16.8 | 18.5 | 20.6 | 40.3 | 43.8 | 47.0 | 50.9 | 53.7 |
| 40 | 20.7 | 22.2 | 24.4 | 26.5 | 29.1 | 51.8 | 55.8 | 59.3 | 63.7 | 66.8 |
| 50 | 28.0 | 29.7 | 32.4 | 34.8 | 37.7 | 63.2 | 67.5 | 71.4 | 76.2 | 79.5 |
| 60 | 35.5 | 37.5 | 40.5 | 43.2 | 46.5 | 74.4 | 79.1 | 83.3 | 88.4 | 92.0 |
| 70 | 43.3 | 45.4 | 48.8 | 51.7 | 55.3 | 85.5 | 90.5 | 95.0 | 100 | 104 |
| 80 | 51.2 | 53.5 | 57.2 | 60.4 | 64.3 | 96.6 | 102 | 107 | 112 | 116 |
| 90 | 59.2 | 61.8 | 65.6 | 69.1 | 73.3 | 108 | 113 | 118 | 124 | 128 |
| 100 | 67.3 | 70.1 | 74.2 | 77.9 | 82.4 | 118 | 124 | 130 | 136 | 140 |

Table T-3: Quantiles, $t_q(\nu)$, for Student's t -distribution with ν degrees of freedom (see Section 7.11)

| ν | q | | | | | | | | | |
|----------|-------|-------|-------|-------|-------|------|-------|-------|------|-------|
| | 0.50 | 0.60 | 0.70 | 0.75 | 0.80 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
| 1 | 0.000 | 0.325 | 0.727 | 1.00 | 1.38 | 3.08 | 6.31 | 12.7 | 31.8 | 63.7 |
| 2 | 0.000 | 0.289 | 0.617 | 0.816 | 1.06 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 |
| 3 | 0.000 | 0.277 | 0.584 | 0.765 | 0.978 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 0.000 | 0.271 | 0.569 | 0.741 | 0.941 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 |
| 5 | 0.000 | 0.267 | 0.559 | 0.727 | 0.920 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 |
| 6 | 0.000 | 0.265 | 0.553 | 0.718 | 0.906 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 |
| 7 | 0.000 | 0.263 | 0.549 | 0.711 | 0.896 | 1.41 | 1.89 | 2.36 | 3.00 | 3.50 |
| 8 | 0.000 | 0.262 | 0.546 | 0.706 | 0.889 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 |
| 9 | 0.000 | 0.261 | 0.543 | 0.703 | 0.883 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 |
| 10 | 0.000 | 0.260 | 0.542 | 0.700 | 0.879 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 |
| 11 | 0.000 | 0.260 | 0.540 | 0.697 | 0.876 | 1.36 | 1.80 | 2.20 | 2.72 | 3.11 |
| 12 | 0.000 | 0.259 | 0.539 | 0.695 | 0.873 | 1.36 | 1.78 | 2.18 | 2.68 | 3.05 |
| 13 | 0.000 | 0.259 | 0.538 | 0.694 | 0.870 | 1.35 | 1.77 | 2.16 | 2.65 | 3.01 |
| 14 | 0.000 | 0.258 | 0.537 | 0.692 | 0.868 | 1.35 | 1.76 | 2.14 | 2.62 | 2.98 |
| 15 | 0.000 | 0.258 | 0.536 | 0.691 | 0.866 | 1.34 | 1.75 | 2.13 | 2.60 | 2.95 |
| 16 | 0.000 | 0.258 | 0.535 | 0.690 | 0.865 | 1.34 | 1.75 | 2.12 | 2.58 | 2.92 |
| 17 | 0.000 | 0.257 | 0.534 | 0.689 | 0.863 | 1.33 | 1.74 | 2.11 | 2.57 | 2.90 |
| 18 | 0.000 | 0.257 | 0.534 | 0.688 | 0.862 | 1.33 | 1.73 | 2.10 | 2.55 | 2.88 |
| 19 | 0.000 | 0.257 | 0.533 | 0.688 | 0.861 | 1.33 | 1.73 | 2.09 | 2.54 | 2.86 |
| 20 | 0.000 | 0.257 | 0.533 | 0.687 | 0.860 | 1.33 | 1.72 | 2.09 | 2.53 | 2.85 |
| 21 | 0.000 | 0.257 | 0.532 | 0.686 | 0.859 | 1.32 | 1.72 | 2.08 | 2.52 | 2.83 |
| 22 | 0.000 | 0.256 | 0.532 | 0.686 | 0.858 | 1.32 | 1.72 | 2.07 | 2.51 | 2.82 |
| 23 | 0.000 | 0.256 | 0.532 | 0.685 | 0.858 | 1.32 | 1.71 | 2.07 | 2.50 | 2.81 |
| 24 | 0.000 | 0.256 | 0.531 | 0.685 | 0.857 | 1.32 | 1.71 | 2.06 | 2.49 | 2.80 |
| 25 | 0.000 | 0.256 | 0.531 | 0.684 | 0.856 | 1.32 | 1.71 | 2.06 | 2.49 | 2.79 |
| 26 | 0.000 | 0.256 | 0.531 | 0.684 | 0.856 | 1.31 | 1.71 | 2.06 | 2.48 | 2.78 |
| 27 | 0.000 | 0.256 | 0.531 | 0.684 | 0.855 | 1.31 | 1.70 | 2.05 | 2.47 | 2.77 |
| 28 | 0.000 | 0.256 | 0.530 | 0.683 | 0.855 | 1.31 | 1.70 | 2.05 | 2.47 | 2.76 |
| 29 | 0.000 | 0.256 | 0.530 | 0.683 | 0.854 | 1.31 | 1.70 | 2.05 | 2.46 | 2.76 |
| 30 | 0.000 | 0.256 | 0.530 | 0.683 | 0.854 | 1.31 | 1.70 | 2.04 | 2.46 | 2.75 |
| 40 | 0.000 | 0.255 | 0.529 | 0.681 | 0.851 | 1.30 | 1.68 | 2.02 | 2.42 | 2.70 |
| 60 | 0.000 | 0.254 | 0.527 | 0.679 | 0.848 | 1.30 | 1.67 | 2.00 | 2.39 | 2.66 |
| ∞ | 0.000 | 0.253 | 0.524 | 0.674 | 0.842 | 1.28 | 1.645 | 1.96 | 2.33 | 2.58 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator (see Section 7.12)

| v_2 | q | v_1 | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.90 | 39.9 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 | 58.9 | 59.4 | 59.9 |
| | 0.95 | 161 | 199 | 216 | 225 | 230 | 234 | 237 | 239 | 241 |
| | 0.975 | 648 | 799 | 864 | 900 | 922 | 937 | 948 | 957 | 963 |
| | 0.99 | 4050 | 5000 | 5400 | 5620 | 5760 | 5860 | 5930 | 5980 | 6020 |
| | 0.995 | 16200 | 20000 | 21600 | 22500 | 23100 | 23400 | 23700 | 23900 | 24100 |
| 2 | 0.90 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 |
| | 0.95 | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 |
| | 0.975 | 38.5 | 39.0 | 39.2 | 39.2 | 39.3 | 39.3 | 39.4 | 39.4 | 39.4 |
| | 0.99 | 98.5 | 99.0 | 99.2 | 99.3 | 99.3 | 99.3 | 99.4 | 99.4 | 99.4 |
| | 0.995 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 |
| 3 | 0.90 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 |
| | 0.95 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 |
| | 0.975 | 17.4 | 16.0 | 15.4 | 15.1 | 14.9 | 14.7 | 14.6 | 14.5 | 14.5 |
| | 0.99 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 | 27.5 | 27.3 |
| | 0.995 | 55.6 | 49.8 | 47.5 | 46.2 | 45.4 | 44.8 | 44.4 | 44.1 | 43.9 |
| 4 | 0.90 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 |
| | 0.95 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 |
| | 0.975 | 12.2 | 10.6 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.90 |
| | 0.99 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 | 14.8 | 14.7 |
| | 0.995 | 31.3 | 26.3 | 24.3 | 23.2 | 22.5 | 22.0 | 21.6 | 21.4 | 21.1 |
| 5 | 0.90 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 |
| | 0.95 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 |
| | 0.975 | 10.0 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.68 |
| | 0.99 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 | 10.3 | 10.2 |
| | 0.995 | 22.8 | 18.3 | 16.5 | 15.6 | 14.9 | 14.5 | 14.2 | 14.0 | 13.8 |
| 6 | 0.90 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 |
| | 0.95 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 |
| | 0.975 | 8.81 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.52 |
| | 0.99 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 |
| | 0.995 | 18.6 | 14.5 | 12.9 | 12.0 | 11.5 | 11.1 | 10.8 | 10.6 | 10.4 |
| 7 | 0.90 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 |
| | 0.95 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 |
| | 0.975 | 8.07 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.82 |
| | 0.99 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 |
| | 0.995 | 16.2 | 12.4 | 10.9 | 10.1 | 9.52 | 9.16 | 8.89 | 8.68 | 8.51 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator
(Continued)

| v_2 | q | v_1 | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| | | 10 | 12 | 15 | 20 | 24 | 30 | 60 | 120 | ∞ |
| 1 | 0.90 | 60.2 | 60.7 | 61.2 | 61.7 | 62.0 | 62.3 | 62.8 | 63.1 | 63.3 |
| | 0.95 | 242 | 244 | 246 | 248 | 249 | 250 | 252 | 253 | 254 |
| | 0.975 | 969 | 977 | 985 | 993 | 997 | 1000 | 1010 | 1010 | 1020 |
| | 0.99 | 6060 | 6110 | 6160 | 6210 | 6230 | 6260 | 6310 | 6340 | 6370 |
| | 0.995 | 24200 | 24400 | 24600 | 24800 | 24900 | 25000 | 25300 | 25400 | 25500 |
| 2 | 0.90 | 9.39 | 9.41 | 9.42 | 9.44 | 9.45 | 9.46 | 9.47 | 9.48 | 9.49 |
| | 0.95 | 19.4 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 |
| | 0.975 | 39.4 | 39.4 | 39.4 | 39.4 | 39.5 | 39.5 | 39.5 | 39.5 | 39.5 |
| | 0.99 | 99.4 | 99.4 | 99.4 | 99.4 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 |
| | 0.995 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 199 | 200 |
| 3 | 0.90 | 5.23 | 5.22 | 5.20 | 5.18 | 5.18 | 5.17 | 5.15 | 5.14 | 5.13 |
| | 0.95 | 8.79 | 8.74 | 8.70 | 8.66 | 8.64 | 8.62 | 8.57 | 8.55 | 8.53 |
| | 0.975 | 14.4 | 14.3 | 14.3 | 14.2 | 14.1 | 14.1 | 14.0 | 13.9 | 13.9 |
| | 0.99 | 27.2 | 27.1 | 26.9 | 26.7 | 26.6 | 26.5 | 26.3 | 26.2 | 26.1 |
| | 0.995 | 43.7 | 43.4 | 43.1 | 42.8 | 42.6 | 42.5 | 42.1 | 42.0 | 41.8 |
| 4 | 0.90 | 3.92 | 3.90 | 3.87 | 3.84 | 3.83 | 3.82 | 3.79 | 3.78 | 3.76 |
| | 0.95 | 5.96 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.69 | 5.66 | 5.63 |
| | 0.975 | 8.84 | 8.75 | 8.66 | 8.56 | 8.51 | 8.46 | 8.36 | 8.31 | 8.26 |
| | 0.99 | 14.5 | 14.4 | 14.2 | 14.0 | 13.9 | 13.8 | 13.7 | 13.6 | 13.5 |
| | 0.995 | 21.0 | 20.7 | 20.4 | 20.2 | 20.0 | 19.9 | 19.6 | 19.5 | 19.3 |
| 5 | 0.90 | 3.30 | 3.27 | 3.24 | 3.21 | 3.19 | 3.17 | 3.14 | 3.12 | 3.11 |
| | 0.95 | 4.74 | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.43 | 4.40 | 4.37 |
| | 0.975 | 6.62 | 6.52 | 6.43 | 6.33 | 6.28 | 6.23 | 6.12 | 6.07 | 6.02 |
| | 0.99 | 10.1 | 9.89 | 9.72 | 9.55 | 9.47 | 9.38 | 9.20 | 9.11 | 9.02 |
| | 0.995 | 13.6 | 13.4 | 13.1 | 12.9 | 12.8 | 12.7 | 12.4 | 12.3 | 12.1 |
| 6 | 0.90 | 2.94 | 2.90 | 2.87 | 2.84 | 2.82 | 2.80 | 2.76 | 2.74 | 2.72 |
| | 0.95 | 4.06 | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.74 | 3.70 | 3.67 |
| | 0.975 | 5.46 | 5.37 | 5.27 | 5.17 | 5.12 | 5.07 | 4.96 | 4.90 | 4.85 |
| | 0.99 | 7.87 | 7.72 | 7.56 | 7.40 | 7.31 | 7.23 | 7.06 | 6.97 | 6.88 |
| | 0.995 | 10.3 | 10.0 | 9.81 | 9.59 | 9.47 | 9.36 | 9.12 | 9.00 | 8.88 |
| 7 | 0.90 | 2.70 | 2.67 | 2.63 | 2.59 | 2.58 | 2.56 | 2.51 | 2.49 | 2.47 |
| | 0.95 | 3.64 | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.30 | 3.27 | 3.23 |
| | 0.975 | 4.76 | 4.67 | 4.57 | 4.47 | 4.41 | 4.36 | 4.25 | 4.20 | 4.14 |
| | 0.99 | 6.62 | 6.47 | 6.31 | 6.16 | 6.07 | 5.99 | 5.82 | 5.74 | 5.65 |
| | 0.995 | 8.38 | 8.18 | 7.97 | 7.75 | 7.64 | 7.53 | 7.31 | 7.19 | 7.08 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator
(Continued)

| v_2 | q | v_1 | | | | | | | | |
|-----------|--------------|-------|-------|------|------|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 8 | 0.90 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 |
| | 0.95 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 |
| | 0.975 | 7.57 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.36 |
| | 0.99 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 |
| | 0.995 | 14.7 | 11.0 | 9.60 | 8.81 | 8.30 | 7.95 | 7.69 | 7.50 | 7.34 |
| 9 | 0.90 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 |
| | 0.95 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 |
| | 0.975 | 7.21 | 5.71 | 5.08 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 4.03 |
| | 0.99 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 |
| | 0.995 | 13.6 | 10.11 | 8.72 | 7.96 | 7.47 | 7.13 | 6.88 | 6.69 | 6.54 |
| 10 | 0.90 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 |
| | 0.95 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 |
| | 0.975 | 6.94 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.78 |
| | 0.99 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 |
| | 0.995 | 12.8 | 9.43 | 8.08 | 7.34 | 6.87 | 6.54 | 6.30 | 6.12 | 5.97 |
| 12 | 0.90 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.21 |
| | 0.95 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 |
| | 0.975 | 6.55 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.44 |
| | 0.99 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 |
| | 0.995 | 11.8 | 8.51 | 7.23 | 6.52 | 6.07 | 5.76 | 5.52 | 5.35 | 5.20 |
| 15 | 0.90 | 3.07 | 2.70 | 2.49 | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.09 |
| | 0.95 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 |
| | 0.975 | 6.20 | 4.77 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 |
| | 0.99 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 |
| | 0.995 | 10.8 | 7.70 | 6.48 | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 |
| 20 | 0.90 | 2.97 | 2.59 | 2.38 | 2.25 | 2.16 | 2.09 | 2.04 | 2.00 | 1.96 |
| | 0.95 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 |
| | 0.975 | 5.87 | 4.46 | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 |
| | 0.99 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 |
| | 0.995 | 9.9 | 6.99 | 5.82 | 5.17 | 4.76 | 4.47 | 4.26 | 4.09 | 3.96 |
| 24 | 0.90 | 2.93 | 2.54 | 2.33 | 2.19 | 2.10 | 2.04 | 1.98 | 1.94 | 1.91 |
| | 0.95 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 |
| | 0.975 | 5.72 | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 |
| | 0.99 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 |
| | 0.995 | 9.6 | 6.66 | 5.52 | 4.89 | 4.49 | 4.20 | 3.99 | 3.83 | 3.69 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator
(Continued)

| v_2 | q | v_1 | | | | | | | | |
|-----------|--------------|-------|------|------|------|------|------|------|------|----------|
| | | 10 | 12 | 15 | 20 | 24 | 30 | 60 | 120 | ∞ |
| 8 | 0.90 | 2.42 | 2.38 | 2.34 | 2.30 | 2.28 | 2.25 | 2.21 | 2.18 | 2.16 |
| | 0.95 | 3.14 | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.79 | 2.75 | 2.71 |
| | 0.975 | 3.96 | 3.87 | 3.77 | 3.67 | 3.61 | 3.56 | 3.45 | 3.39 | 3.33 |
| | 0.99 | 5.3 | 5.11 | 4.96 | 4.81 | 4.73 | 4.65 | 4.48 | 4.40 | 4.31 |
| | 0.995 | 6.4 | 6.2 | 6.03 | 5.83 | 5.73 | 5.62 | 5.41 | 5.30 | 5.19 |
| 9 | 0.90 | 2.32 | 2.28 | 2.24 | 2.20 | 2.18 | 2.16 | 2.11 | 2.08 | 2.06 |
| | 0.95 | 2.98 | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.62 | 2.58 | 2.54 |
| | 0.975 | 3.72 | 3.62 | 3.52 | 3.42 | 3.37 | 3.31 | 3.20 | 3.14 | 3.08 |
| | 0.99 | 4.8 | 4.71 | 4.56 | 4.41 | 4.33 | 4.25 | 4.08 | 4.00 | 3.91 |
| | 0.995 | 5.8 | 5.66 | 5.47 | 5.27 | 5.17 | 5.07 | 4.86 | 4.75 | 4.64 |
| 10 | 0.90 | 2.19 | 2.15 | 2.10 | 2.06 | 2.04 | 2.01 | 1.96 | 1.93 | 1.90 |
| | 0.95 | 2.75 | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.38 | 2.34 | 2.30 |
| | 0.975 | 3.37 | 3.28 | 3.18 | 3.07 | 3.02 | 2.96 | 2.85 | 2.79 | 2.73 |
| | 0.99 | 4.30 | 4.16 | 4.01 | 3.86 | 3.78 | 3.70 | 3.54 | 3.45 | 3.36 |
| | 0.995 | 5.1 | 4.91 | 4.72 | 4.53 | 4.43 | 4.33 | 4.12 | 4.01 | 3.90 |
| 12 | 0.90 | 2.06 | 2.02 | 1.97 | 1.92 | 1.90 | 1.87 | 1.82 | 1.79 | 1.76 |
| | 0.95 | 2.54 | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.16 | 2.11 | 2.07 |
| | 0.975 | 3.06 | 2.96 | 2.86 | 2.76 | 2.70 | 2.64 | 2.52 | 2.46 | 2.40 |
| | 0.99 | 3.80 | 3.67 | 3.52 | 3.37 | 3.29 | 3.21 | 3.05 | 2.96 | 2.87 |
| | 0.995 | 4.4 | 4.25 | 4.07 | 3.88 | 3.79 | 3.69 | 3.48 | 3.37 | 3.26 |
| 15 | 0.90 | 1.94 | 1.89 | 1.84 | 1.79 | 1.77 | 1.74 | 1.68 | 1.64 | 1.61 |
| | 0.95 | 2.35 | 2.28 | 2.20 | 2.12 | 2.08 | 2.04 | 1.95 | 1.90 | 1.84 |
| | 0.975 | 2.77 | 2.68 | 2.57 | 2.46 | 2.41 | 2.35 | 2.22 | 2.16 | 2.09 |
| | 0.99 | 3.37 | 3.23 | 3.09 | 2.94 | 2.86 | 2.78 | 2.61 | 2.52 | 2.42 |
| | 0.995 | 3.8 | 3.68 | 3.50 | 3.32 | 3.22 | 3.12 | 2.92 | 2.81 | 2.69 |
| 20 | 0.90 | 1.88 | 1.83 | 1.78 | 1.73 | 1.70 | 1.67 | 1.61 | 1.57 | 1.53 |
| | 0.95 | 2.25 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.84 | 1.79 | 1.73 |
| | 0.975 | 2.64 | 2.54 | 2.44 | 2.33 | 2.27 | 2.21 | 2.08 | 2.01 | 1.94 |
| | 0.99 | 3.17 | 3.03 | 2.89 | 2.74 | 2.66 | 2.58 | 2.40 | 2.31 | 2.21 |
| | 0.995 | 3.6 | 3.42 | 3.25 | 3.06 | 2.97 | 2.87 | 2.66 | 2.55 | 2.43 |
| 24 | 0.90 | 1.88 | 1.83 | 1.78 | 1.73 | 1.70 | 1.67 | 1.61 | 1.57 | 1.53 |
| | 0.95 | 2.25 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.84 | 1.79 | 1.73 |
| | 0.975 | 2.64 | 2.54 | 2.44 | 2.33 | 2.27 | 2.21 | 2.08 | 2.01 | 1.94 |
| | 0.99 | 3.17 | 3.03 | 2.89 | 2.74 | 2.66 | 2.58 | 2.40 | 2.31 | 2.21 |
| | 0.995 | 3.6 | 3.42 | 3.25 | 3.06 | 2.97 | 2.87 | 2.66 | 2.55 | 2.43 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator (continued)

| v_2 | q | v_1 | | | | | | | | |
|----------|-------|-------|------|------|------|------|------|------|------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 30 | 0.90 | 2.88 | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.93 | 1.88 | 1.85 |
| | 0.95 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 |
| | 0.975 | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 |
| | 0.99 | 7.6 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 |
| | 0.995 | 9.2 | 6.4 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 |
| 60 | 0.90 | 2.79 | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.82 | 1.77 | 1.74 |
| | 0.95 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 |
| | 0.975 | 5.29 | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 |
| | 0.99 | 7.1 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 |
| | 0.995 | 8.5 | 5.8 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 |
| 120 | 0.90 | 2.75 | 2.35 | 2.13 | 1.99 | 1.90 | 1.82 | 1.77 | 1.72 | 1.68 |
| | 0.95 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 |
| | 0.975 | 5.15 | 3.80 | 3.23 | 2.89 | 2.67 | 2.52 | 2.39 | 2.30 | 2.22 |
| | 0.99 | 6.9 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 |
| | 0.995 | 8.2 | 5.54 | 4.50 | 3.92 | 3.55 | 3.28 | 3.09 | 2.93 | 2.81 |
| ∞ | 0.90 | 2.71 | 2.30 | 2.08 | 1.94 | 1.85 | 1.77 | 1.72 | 1.67 | 1.63 |
| | 0.95 | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 |
| | 0.975 | 5.02 | 3.69 | 3.12 | 2.79 | 2.57 | 2.41 | 2.29 | 2.19 | 2.11 |
| | 0.99 | 6.64 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 |
| | 0.995 | 7.9 | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 |

Table T-4: Quantiles, $f_q(v_1, v_2)$, for the F -distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator
(continued)

| v_2 | q | v_1 | | | | | | | | |
|------------|--------------|-------|------|------|------|------|------|------|------|----------|
| | | 10 | 12 | 15 | 20 | 24 | 30 | 60 | 120 | ∞ |
| 30 | 0.90 | 1.82 | 1.77 | 1.72 | 1.67 | 1.64 | 1.61 | 1.54 | 1.50 | 1.46 |
| | 0.95 | 2.16 | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.74 | 1.68 | 1.62 |
| | 0.975 | 2.51 | 2.41 | 2.31 | 2.20 | 2.14 | 2.07 | 1.94 | 1.87 | 1.79 |
| | 0.99 | 2.98 | 2.84 | 2.70 | 2.55 | 2.47 | 2.39 | 2.21 | 2.11 | 2.01 |
| | 0.995 | 3.34 | 3.18 | 3.01 | 2.82 | 2.73 | 2.63 | 2.42 | 2.30 | 2.18 |
| 60 | 0.90 | 1.71 | 1.66 | 1.60 | 1.54 | 1.51 | 1.48 | 1.40 | 1.35 | 1.29 |
| | 0.95 | 1.99 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.53 | 1.47 | 1.39 |
| | 0.975 | 2.27 | 2.17 | 2.06 | 1.94 | 1.88 | 1.82 | 1.67 | 1.58 | 1.48 |
| | 0.99 | 2.63 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.84 | 1.73 | 1.60 |
| | 0.995 | 2.90 | 2.74 | 2.57 | 2.39 | 2.29 | 2.19 | 1.96 | 1.83 | 1.69 |
| 120 | 0.90 | 1.65 | 1.60 | 1.55 | 1.48 | 1.45 | 1.41 | 1.32 | 1.26 | 1.19 |
| | 0.95 | 1.91 | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.43 | 1.35 | 1.25 |
| | 0.975 | 2.16 | 2.05 | 1.94 | 1.82 | 1.76 | 1.69 | 1.53 | 1.43 | 1.31 |
| | 0.99 | 2.47 | 2.34 | 2.19 | 2.03 | 1.95 | 1.86 | 1.66 | 1.53 | 1.38 |
| | 0.995 | 2.71 | 2.54 | 2.37 | 2.19 | 2.09 | 1.98 | 1.75 | 1.61 | 1.43 |
| ∞ | 0.90 | 1.60 | 1.55 | 1.49 | 1.42 | 1.38 | 1.34 | 1.24 | 1.17 | 1.00 |
| | 0.95 | 1.83 | 1.75 | 1.67 | 1.57 | 1.52 | 1.46 | 1.32 | 1.22 | 1.00 |
| | 0.975 | 2.05 | 1.94 | 1.83 | 1.71 | 1.64 | 1.57 | 1.39 | 1.27 | 1.00 |
| | 0.99 | 2.32 | 2.18 | 2.04 | 1.88 | 1.79 | 1.70 | 1.47 | 1.32 | 1.00 |
| | 0.995 | 2.52 | 2.36 | 2.19 | 2.00 | 1.90 | 1.79 | 1.53 | 1.36 | 1.00 |

Table T-5: Quantiles $f_{max,q}(k, v)$ for the F_{max} distribution (see Section 14.5)

| v | $k = \text{number of groups}$ | | | | | | | | | | | q |
|----------|-------------------------------|------|------|------|------|------|------|------|------|------|------|-----|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| 2 | 39.0 | 87.5 | 142 | 202 | 266 | 333 | 403 | 475 | 550 | 626 | 704 | |
| 3 | 15.4 | 27.8 | 39.2 | 50.7 | 62.0 | 72.9 | 83.5 | 93.9 | 104 | 114 | 124 | |
| 4 | 9.60 | 15.5 | 20.6 | 25.2 | 29.5 | 33.6 | 37.5 | 41.1 | 44.6 | 48.0 | 51.4 | |
| 5 | 7.15 | 10.8 | 13.7 | 16.3 | 18.7 | 20.8 | 22.9 | 24.7 | 26.5 | 28.2 | 29.9 | |
| 6 | 5.82 | 8.38 | 10.4 | 12.1 | 13.7 | 15.0 | 16.3 | 17.5 | 18.6 | 19.7 | 20.7 | |
| 7 | 4.99 | 6.94 | 8.44 | 9.70 | 10.8 | 11.8 | 12.7 | 13.5 | 14.3 | 15.1 | 15.8 | |
| 8 | 4.43 | 6.00 | 7.18 | 8.12 | 9.03 | 9.78 | 10.5 | 11.1 | 11.7 | 12.2 | 12.7 | |
| 9 | 4.03 | 5.34 | 6.31 | 7.11 | 7.80 | 8.41 | 8.95 | 9.45 | 9.91 | 10.3 | 10.7 | |
| 10 | 3.72 | 4.85 | 5.67 | 6.34 | 6.92 | 7.42 | 7.87 | 8.28 | 8.66 | 9.01 | 9.34 | |
| 12 | 3.28 | 4.16 | 4.79 | 5.30 | 5.72 | 6.09 | 6.42 | 6.72 | 7.00 | 7.25 | 7.48 | |
| 15 | 2.86 | 3.54 | 4.01 | 4.37 | 4.68 | 4.95 | 5.19 | 5.40 | 5.59 | 5.77 | 5.93 | |
| 20 | 2.46 | 2.95 | 3.29 | 3.54 | 3.76 | 3.94 | 4.10 | 4.24 | 4.37 | 4.49 | 4.59 | |
| 30 | 2.07 | 2.40 | 2.61 | 2.78 | 2.91 | 3.02 | 3.12 | 3.21 | 3.29 | 3.36 | 3.39 | |
| 60 | 1.67 | 1.85 | 1.96 | 2.04 | 2.11 | 2.17 | 2.22 | 2.26 | 2.30 | 2.33 | 2.36 | |
| ∞ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |

| v | $k = \text{number of groups}$ | | | | | | | | | | | $q = 0.99$ |
|----------|-------------------------------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|------------|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| 2 | 199 | 448 | 729 | 1,036 | 1,362 | 1,705 | 2,063 | 2,432 | 2,813 | 3,204 | 3,605 | |
| 3 | 47.5 | 85 | 120 | 151 | 184 | 216 | 249 | 281 | 310 | 337 | 361 | |
| 4 | 23.2 | 37 | 49 | 59 | 69 | 79 | 89 | 97 | 106 | 113 | 120 | |
| 5 | 14.9 | 22 | 28 | 33 | 38 | 42 | 46 | 50 | 54 | 57 | 60 | |
| 6 | 11.1 | 15.5 | 19.1 | 22 | 25 | 27 | 30 | 32 | 34 | 36 | 37 | |
| 7 | 8.89 | 12.1 | 14.5 | 16.5 | 18.4 | 20 | 22 | 23 | 24 | 26 | 27 | |
| 8 | 7.50 | 9.9 | 11.7 | 13.2 | 14.5 | 15.8 | 16.9 | 17.9 | 18.9 | 19.8 | 21 | |
| 9 | 6.54 | 8.5 | 9.9 | 11.1 | 12.1 | 13.1 | 13.9 | 14.7 | 15.3 | 16.0 | 16.6 | |
| 10 | 5.85 | 7.4 | 8.6 | 9.6 | 10.4 | 11.1 | 11.8 | 12.4 | 12.9 | 13.4 | 13.9 | |
| 12 | 4.91 | 6.1 | 6.9 | 7.6 | 8.2 | 8.7 | 9.1 | 9.5 | 9.9 | 10.2 | 10.6 | |
| 15 | 4.07 | 4.9 | 5.5 | 6.0 | 6.4 | 6.7 | 7.1 | 7.3 | 7.5 | 7.8 | 8.0 | |
| 20 | 3.32 | 3.8 | 4.3 | 4.6 | 4.9 | 5.1 | 5.3 | 5.5 | 5.6 | 5.8 | 5.9 | |
| 30 | 2.63 | 3.0 | 3.3 | 3.4 | 3.6 | 3.7 | 3.8 | 3.9 | 4.0 | 4.1 | 4.2 | |
| 60 | 1.96 | 2.2 | 2.3 | 2.4 | 2.4 | 2.5 | 2.5 | 2.6 | 2.6 | 2.7 | 2.7 | |
| ∞ | 1.00 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | |

Adapted from David, H. A., 1952, Upper 5 and 1% Points of the Maximum F-Ratio, *Biometrika*, **39**, pp. 422-424, with permission of the Biometrika Trustees.

Table T-6a: Coefficients $\{a_{n-i+1}\}$ for the W - test for normality
(see Section 11.9)

| $i \setminus n$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | |
|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0.7071 | 0.6872 | 0.6646 | 0.6431 | 0.6233 | 0.6052 | 0.5888 | 0.5739 | | |
| 2 | | 0.1677 | 0.2413 | 0.2806 | 0.3031 | 0.3164 | 0.3244 | 0.3291 | | |
| 3 | | | | 0.0875 | 0.1401 | 0.1743 | 0.1976 | 0.2141 | | |
| 4 | | | | | | 0.0561 | 0.0947 | 0.1224 | | |
| 5 | | | | | | | | 0.0399 | | |
| $i \setminus n$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | 0.5601 | 0.5475 | 0.5359 | 0.5251 | 0.5150 | 0.5056 | 0.4968 | 0.4886 | 0.4808 | 0.4734 |
| 2 | 0.3315 | 0.3325 | 0.3325 | 0.3318 | 0.3306 | 0.3290 | 0.3273 | 0.3253 | 0.3232 | 0.3211 |
| 3 | 0.2260 | 0.2347 | 0.2412 | 0.2460 | 0.2495 | 0.2521 | 0.2540 | 0.2553 | 0.2561 | 0.2565 |
| 4 | 0.1429 | 0.1586 | 0.1707 | 0.1802 | 0.1878 | 0.1939 | 0.1988 | 0.2027 | 0.2059 | 0.2085 |
| 5 | 0.0695 | 0.0922 | 0.1099 | 0.1240 | 0.1353 | 0.1447 | 0.1524 | 0.1587 | 0.1641 | 0.1686 |
| 6 | | 0.0303 | 0.0539 | 0.0727 | 0.0880 | 0.1005 | 0.1109 | 0.1197 | 0.1271 | 0.1334 |
| 7 | | | 0.0240 | 0.0433 | 0.0593 | 0.0725 | 0.0837 | 0.0932 | 0.1013 | |
| 8 | | | | | 0.0196 | 0.0359 | 0.0496 | 0.0612 | 0.0711 | |
| 9 | | | | | | | 0.0163 | 0.0303 | 0.0422 | |
| 10 | | | | | | | | | | 0.0140 |
| $i \setminus n$ | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 1 | 0.4643 | 0.4590 | 0.4542 | 0.4493 | 0.4450 | 0.4407 | 0.4366 | 0.4328 | 0.4291 | 0.4254 |
| 2 | 0.3185 | 0.3156 | 0.3126 | 0.3098 | 0.3069 | 0.3043 | 0.3018 | 0.2992 | 0.2968 | 0.2944 |
| 3 | 0.2578 | 0.2571 | 0.2563 | 0.2554 | 0.2543 | 0.2533 | 0.2522 | 0.2510 | 0.2499 | 0.2487 |
| 4 | 0.2119 | 0.2131 | 0.2139 | 0.2145 | 0.2148 | 0.2151 | 0.2152 | 0.2151 | 0.2150 | 0.2148 |
| 5 | 0.1736 | 0.1764 | 0.1787 | 0.1807 | 0.1822 | 0.1836 | 0.1848 | 0.1857 | 0.1864 | 0.1870 |
| 6 | 0.1399 | 0.1443 | 0.1480 | 0.1512 | 0.1539 | 0.1563 | 0.1584 | 0.1601 | 0.1616 | 0.1630 |
| 7 | 0.1092 | 0.1150 | 0.1201 | 0.1245 | 0.1283 | 0.1316 | 0.1346 | 0.1372 | 0.1395 | 0.1415 |
| 8 | 0.0804 | 0.0878 | 0.0941 | 0.0997 | 0.1046 | 0.1089 | 0.1128 | 0.1162 | 0.1192 | 0.1219 |
| 9 | 0.0530 | 0.0618 | 0.0696 | 0.0764 | 0.0823 | 0.0876 | 0.0923 | 0.0965 | 0.1002 | 0.1036 |
| 10 | 0.0263 | 0.0368 | 0.0459 | 0.0539 | 0.0610 | 0.0672 | 0.0728 | 0.0778 | 0.0822 | 0.0862 |
| 11 | | 0.0122 | 0.0228 | 0.0321 | 0.0403 | 0.0476 | 0.0540 | 0.0598 | 0.0650 | 0.0697 |
| 12 | | | 0.0107 | 0.0200 | 0.0284 | 0.0358 | 0.0424 | 0.0483 | 0.0537 | |
| 13 | | | | | 0.0094 | 0.0178 | 0.0253 | 0.0320 | 0.0381 | |
| 14 | | | | | | | 0.0084 | 0.0159 | 0.0227 | |
| 15 | | | | | | | | | | 0.0076 |

Adapted from Shapiro, S. S., and M. B. Wilk, 1965, An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, 52, pp. 591-611, with permission of the Biometrika Trustees.

Table T-6b: Quantiles, $w_q(n)$, for the W -test for normality (see Section 11.9)

| n | $q = 0.01$ | $q = 0.02$ | $q = 0.05$ | $q = 0.10$ | $q = 0.50$ |
|-----|------------|------------|------------|------------|------------|
| 3 | 0.753 | 0.756 | 0.767 | 0.789 | 0.959 |
| 4 | 0.687 | 0.707 | 0.748 | 0.792 | 0.935 |
| 5 | 0.686 | 0.715 | 0.762 | 0.806 | 0.927 |
| 6 | 0.713 | 0.743 | 0.788 | 0.826 | 0.927 |
| 7 | 0.730 | 0.760 | 0.803 | 0.838 | 0.928 |
| 8 | 0.749 | 0.778 | 0.818 | 0.851 | 0.932 |
| 9 | 0.764 | 0.791 | 0.829 | 0.859 | 0.935 |
| 10 | 0.781 | 0.806 | 0.842 | 0.869 | 0.938 |
| 11 | 0.792 | 0.817 | 0.850 | 0.876 | 0.940 |
| 12 | 0.805 | 0.828 | 0.859 | 0.883 | 0.943 |
| 13 | 0.814 | 0.837 | 0.866 | 0.889 | 0.945 |
| 14 | 0.825 | 0.846 | 0.874 | 0.895 | 0.947 |
| 15 | 0.835 | 0.855 | 0.881 | 0.901 | 0.950 |
| 16 | 0.844 | 0.863 | 0.887 | 0.906 | 0.952 |
| 17 | 0.851 | 0.869 | 0.892 | 0.910 | 0.954 |
| 18 | 0.858 | 0.874 | 0.897 | 0.914 | 0.956 |
| 19 | 0.863 | 0.879 | 0.901 | 0.917 | 0.957 |
| 20 | 0.868 | 0.884 | 0.905 | 0.920 | 0.959 |
| 21 | 0.873 | 0.888 | 0.908 | 0.923 | 0.960 |
| 22 | 0.878 | 0.892 | 0.911 | 0.926 | 0.961 |
| 23 | 0.881 | 0.895 | 0.914 | 0.928 | 0.962 |
| 24 | 0.884 | 0.898 | 0.916 | 0.930 | 0.963 |
| 25 | 0.888 | 0.901 | 0.918 | 0.931 | 0.964 |
| 26 | 0.891 | 0.904 | 0.920 | 0.933 | 0.965 |
| 27 | 0.894 | 0.906 | 0.923 | 0.935 | 0.965 |
| 28 | 0.896 | 0.908 | 0.924 | 0.936 | 0.966 |
| 29 | 0.898 | 0.910 | 0.926 | 0.937 | 0.966 |
| 30 | 0.900 | 0.912 | 0.927 | 0.939 | 0.967 |
| 32 | 0.904 | 0.915 | 0.930 | 0.941 | 0.968 |
| 34 | 0.908 | 0.919 | 0.933 | 0.943 | 0.969 |
| 36 | 0.912 | 0.922 | 0.935 | 0.945 | 0.970 |
| 38 | 0.916 | 0.925 | 0.938 | 0.947 | 0.971 |
| 40 | 0.919 | 0.928 | 0.940 | 0.949 | 0.972 |
| 42 | 0.922 | 0.930 | 0.942 | 0.951 | 0.972 |
| 44 | 0.924 | 0.933 | 0.944 | 0.952 | 0.973 |
| 46 | 0.927 | 0.935 | 0.945 | 0.953 | 0.974 |
| 48 | 0.929 | 0.937 | 0.947 | 0.954 | 0.974 |
| 50 | 0.930 | 0.938 | 0.947 | 0.955 | 0.974 |

Adapted from Shapiro, S. S., and M. B. Wilk, 1965, An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, **52**, pp. 591-611, with permission of the Biometrika Trustees.

Table T-7: Significant ranges, $q_{0.95}(p, N - k)$, for Duncan's multiple-range test (see Section 16.8)

| $N - k$ | p | | | | | | | | | | | |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 20 | 50 | 100 |
| 1 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 |
| 2 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 |
| 3 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 |
| 4 | 3.93 | 4.01 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 |
| 5 | 3.64 | 3.74 | 3.79 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 |
| 6 | 3.46 | 3.58 | 3.64 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 |
| 7 | 3.35 | 3.47 | 3.54 | 3.58 | 3.60 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 |
| 8 | 3.26 | 3.39 | 3.47 | 3.52 | 3.55 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 |
| 9 | 3.20 | 3.34 | 3.41 | 3.47 | 3.50 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 |
| 10 | 3.15 | 3.30 | 3.37 | 3.43 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 |
| 11 | 3.11 | 3.27 | 3.35 | 3.39 | 3.43 | 3.44 | 3.45 | 3.46 | 3.46 | 3.48 | 3.48 | 3.48 |
| 12 | 3.08 | 3.23 | 3.33 | 3.36 | 3.40 | 3.42 | 3.44 | 3.44 | 3.46 | 3.48 | 3.48 | 3.48 |
| 13 | 3.06 | 3.21 | 3.30 | 3.35 | 3.38 | 3.41 | 3.42 | 3.44 | 3.45 | 3.47 | 3.47 | 3.47 |
| 14 | 3.03 | 3.18 | 3.27 | 3.33 | 3.37 | 3.39 | 3.41 | 3.42 | 3.44 | 3.47 | 3.47 | 3.47 |
| 15 | 3.01 | 3.16 | 3.25 | 3.31 | 3.36 | 3.38 | 3.40 | 3.42 | 3.43 | 3.47 | 3.47 | 3.47 |
| 16 | 3.00 | 3.15 | 3.23 | 3.30 | 3.34 | 3.37 | 3.39 | 3.41 | 3.43 | 3.47 | 3.47 | 3.47 |
| 17 | 2.98 | 3.13 | 3.22 | 3.28 | 3.33 | 3.36 | 3.38 | 3.40 | 3.42 | 3.47 | 3.47 | 3.47 |
| 18 | 2.97 | 3.12 | 3.21 | 3.27 | 3.32 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 19 | 2.96 | 3.11 | 3.19 | 3.26 | 3.31 | 3.35 | 3.37 | 3.39 | 3.41 | 3.47 | 3.47 | 3.47 |
| 20 | 2.95 | 3.10 | 3.18 | 3.25 | 3.30 | 3.34 | 3.36 | 3.38 | 3.40 | 3.47 | 3.47 | 3.47 |
| 30 | 2.89 | 3.04 | 3.12 | 3.20 | 3.25 | 3.29 | 3.32 | 3.35 | 3.37 | 3.47 | 3.47 | 3.47 |
| 40 | 2.86 | 3.01 | 3.10 | 3.17 | 3.22 | 3.27 | 3.30 | 3.33 | 3.35 | 3.47 | 3.47 | 3.47 |
| 60 | 2.83 | 2.98 | 3.08 | 3.14 | 3.20 | 3.24 | 3.28 | 3.31 | 3.33 | 3.47 | 3.48 | 3.48 |
| 100 | 2.80 | 2.95 | 3.05 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.47 | 3.53 | 3.53 |
| ∞ | 2.77 | 2.92 | 3.02 | 3.09 | 3.15 | 3.19 | 3.23 | 3.26 | 3.29 | 3.47 | 3.61 | 3.67 |

Reproduced from: D.B. Duncan, "Multiple Range and Multiple F-tests." BIOMETRICS 11: 1-42. 1955. With permission from The Biometric Society.

Table T-8: Selected binomial probabilities (see Section 22.3)

| n | $y = \text{number of attributes in a binomial sample of size } n \text{ with } \pi = 0.01$ | | | | | | | | | |
|-----|--|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.9900 | 0.0100 | | | | | | | | |
| 2 | 0.9801 | 0.0198 | 0.0001 | | | | | | | |
| 3 | 0.9703 | 0.0294 | 0.0003 | 0.0000 | | | | | | |
| 4 | 0.9606 | 0.0388 | 0.0006 | 0.0000 | 0.0000 | | | | | |
| 5 | 0.9510 | 0.0480 | 0.0010 | 0.0000 | 0.0000 | 0.0000 | | | | |
| 6 | 0.9415 | 0.0571 | 0.0014 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | | | |
| 7 | 0.9321 | 0.0659 | 0.0020 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | | |
| 8 | 0.9227 | 0.0746 | 0.0026 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | |
| 9 | 0.9135 | 0.0830 | 0.0034 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 10 | 0.9044 | 0.0914 | 0.0042 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 11 | 0.8953 | 0.0995 | 0.0050 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 12 | 0.8864 | 0.1074 | 0.0060 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 13 | 0.8775 | 0.1152 | 0.0070 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 14 | 0.8687 | 0.1229 | 0.0081 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 15 | 0.8601 | 0.1303 | 0.0092 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 16 | 0.8515 | 0.1376 | 0.0104 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 17 | 0.8429 | 0.1447 | 0.0117 | 0.0006 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 18 | 0.8345 | 0.1517 | 0.0130 | 0.0007 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 19 | 0.8262 | 0.1586 | 0.0144 | 0.0008 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 20 | 0.8179 | 0.1652 | 0.0159 | 0.0010 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 21 | 0.8097 | 0.1718 | 0.0173 | 0.0011 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 22 | 0.8016 | 0.1781 | 0.0189 | 0.0013 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 23 | 0.7936 | 0.1844 | 0.0205 | 0.0014 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 24 | 0.7857 | 0.1905 | 0.0221 | 0.0016 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 25 | 0.7778 | 0.1964 | 0.0238 | 0.0018 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 26 | 0.7700 | 0.2022 | 0.0255 | 0.0021 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 27 | 0.7623 | 0.2079 | 0.0273 | 0.0023 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 28 | 0.7547 | 0.2135 | 0.0291 | 0.0025 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 29 | 0.7472 | 0.2189 | 0.0310 | 0.0028 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 30 | 0.7397 | 0.2242 | 0.0328 | 0.0031 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 32 | 0.7250 | 0.2343 | 0.0367 | 0.0037 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 34 | 0.7106 | 0.2440 | 0.0407 | 0.0044 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 36 | 0.6964 | 0.2532 | 0.0448 | 0.0051 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 38 | 0.6826 | 0.2620 | 0.0490 | 0.0059 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 40 | 0.6690 | 0.2703 | 0.0532 | 0.0068 | 0.0006 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Table T-8: Selected binomial probabilities (Continued)

| n | y = number of attributes in a binomial sample of size n with $\pi = 0.05$ | | | | | | | | | |
|----|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.9500 | 0.0500 | | | | | | | | |
| 2 | 0.9025 | 0.0950 | 0.0025 | | | | | | | |
| 3 | 0.8574 | 0.1354 | 0.0071 | 0.0001 | | | | | | |
| 4 | 0.8145 | 0.1715 | 0.0135 | 0.0005 | 0.0000 | | | | | |
| 5 | 0.7738 | 0.2036 | 0.0214 | 0.0011 | 0.0000 | 0.0000 | | | | |
| 6 | 0.7351 | 0.2321 | 0.0305 | 0.0021 | 0.0001 | 0.0000 | 0.0000 | | | |
| 7 | 0.6983 | 0.2573 | 0.0406 | 0.0036 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | | |
| 8 | 0.6634 | 0.2793 | 0.0515 | 0.0054 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | |
| 9 | 0.6302 | 0.2985 | 0.0629 | 0.0077 | 0.0006 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 10 | 0.5987 | 0.3151 | 0.0746 | 0.0105 | 0.0010 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 11 | 0.5688 | 0.3293 | 0.0867 | 0.0137 | 0.0014 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 12 | 0.5404 | 0.3413 | 0.0988 | 0.0173 | 0.0021 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 13 | 0.5133 | 0.3512 | 0.1109 | 0.0214 | 0.0028 | 0.0003 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 14 | 0.4877 | 0.3593 | 0.1229 | 0.0259 | 0.0037 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 15 | 0.4633 | 0.3658 | 0.1348 | 0.0307 | 0.0049 | 0.0006 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 16 | 0.4401 | 0.3706 | 0.1463 | 0.0359 | 0.0061 | 0.0008 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 17 | 0.4181 | 0.3741 | 0.1575 | 0.0415 | 0.0076 | 0.0010 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 18 | 0.3972 | 0.3763 | 0.1683 | 0.0473 | 0.0093 | 0.0014 | 0.0002 | 0.0000 | 0.0000 | 0.0000 |
| 19 | 0.3774 | 0.3774 | 0.1787 | 0.0533 | 0.0112 | 0.0018 | 0.0002 | 0.0000 | 0.0000 | 0.0000 |
| 20 | 0.3585 | 0.3774 | 0.1887 | 0.0596 | 0.0133 | 0.0022 | 0.0003 | 0.0000 | 0.0000 | 0.0000 |
| 21 | 0.3406 | 0.3764 | 0.1981 | 0.0660 | 0.0156 | 0.0028 | 0.0004 | 0.0000 | 0.0000 | 0.0000 |
| 22 | 0.3235 | 0.3746 | 0.2070 | 0.0726 | 0.0182 | 0.0034 | 0.0005 | 0.0001 | 0.0000 | 0.0000 |
| 23 | 0.3074 | 0.3721 | 0.2154 | 0.0794 | 0.0209 | 0.0042 | 0.0007 | 0.0001 | 0.0000 | 0.0000 |
| 24 | 0.2920 | 0.3688 | 0.2232 | 0.0862 | 0.0238 | 0.0050 | 0.0008 | 0.0001 | 0.0000 | 0.0000 |
| 25 | 0.2774 | 0.3650 | 0.2305 | 0.0930 | 0.0269 | 0.0060 | 0.0010 | 0.0001 | 0.0000 | 0.0000 |
| 26 | 0.2635 | 0.3606 | 0.2372 | 0.0999 | 0.0302 | 0.0070 | 0.0013 | 0.0002 | 0.0000 | 0.0000 |
| 27 | 0.2503 | 0.3558 | 0.2434 | 0.1068 | 0.0337 | 0.0082 | 0.0016 | 0.0002 | 0.0000 | 0.0000 |
| 28 | 0.2378 | 0.3505 | 0.2490 | 0.1136 | 0.0374 | 0.0094 | 0.0019 | 0.0003 | 0.0000 | 0.0000 |
| 29 | 0.2259 | 0.3448 | 0.2541 | 0.1204 | 0.0412 | 0.0108 | 0.0023 | 0.0004 | 0.0001 | 0.0000 |
| 30 | 0.2146 | 0.3389 | 0.2586 | 0.1270 | 0.0451 | 0.0124 | 0.0027 | 0.0005 | 0.0001 | 0.0000 |
| 32 | 0.1937 | 0.3263 | 0.2662 | 0.1401 | 0.0535 | 0.0158 | 0.0037 | 0.0007 | 0.0001 | 0.0000 |
| 34 | 0.1748 | 0.3128 | 0.2717 | 0.1525 | 0.0622 | 0.0196 | 0.0050 | 0.0011 | 0.0002 | 0.0000 |
| 36 | 0.1578 | 0.2990 | 0.2753 | 0.1642 | 0.0713 | 0.0240 | 0.0065 | 0.0015 | 0.0003 | 0.0000 |
| 38 | 0.1424 | 0.2848 | 0.2773 | 0.1751 | 0.0807 | 0.0289 | 0.0084 | 0.0020 | 0.0004 | 0.0001 |
| 40 | 0.1285 | 0.2706 | 0.2777 | 0.1851 | 0.0901 | 0.0342 | 0.0105 | 0.0027 | 0.0006 | 0.0001 |

Table T-8: Selected binomial probabilities (Continued)

| <i>n</i> | <i>y</i> = number of attributes in a binomial sample of size <i>n</i> with $\pi = 0.10$ | | | | | | | | | |
|----------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.9000 | 0.1000 | | | | | | | | |
| 2 | 0.8100 | 0.1800 | 0.0100 | | | | | | | |
| 3 | 0.7290 | 0.2430 | 0.0270 | 0.0010 | | | | | | |
| 4 | 0.6561 | 0.2916 | 0.0486 | 0.0036 | 0.0001 | | | | | |
| 5 | 0.5905 | 0.3281 | 0.0729 | 0.0081 | 0.0005 | 0.0000 | | | | |
| 6 | 0.5314 | 0.3543 | 0.0984 | 0.0146 | 0.0012 | 0.0001 | 0.0000 | | | |
| 7 | 0.4783 | 0.3720 | 0.1240 | 0.0230 | 0.0026 | 0.0002 | 0.0000 | 0.0000 | | |
| 8 | 0.4305 | 0.3826 | 0.1488 | 0.0331 | 0.0046 | 0.0004 | 0.0000 | 0.0000 | 0.0000 | |
| 9 | 0.3874 | 0.3874 | 0.1722 | 0.0446 | 0.0074 | 0.0008 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 10 | 0.3487 | 0.3874 | 0.1937 | 0.0574 | 0.0112 | 0.0015 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 11 | 0.3138 | 0.3835 | 0.2131 | 0.0710 | 0.0158 | 0.0025 | 0.0003 | 0.0000 | 0.0000 | 0.0000 |
| 12 | 0.2824 | 0.3766 | 0.2301 | 0.0852 | 0.0213 | 0.0038 | 0.0005 | 0.0000 | 0.0000 | 0.0000 |
| 13 | 0.2542 | 0.3672 | 0.2448 | 0.0997 | 0.0277 | 0.0055 | 0.0008 | 0.0001 | 0.0000 | 0.0000 |
| 14 | 0.2288 | 0.3559 | 0.2570 | 0.1142 | 0.0349 | 0.0078 | 0.0013 | 0.0002 | 0.0000 | 0.0000 |
| 15 | 0.2059 | 0.3432 | 0.2669 | 0.1285 | 0.0428 | 0.0105 | 0.0019 | 0.0003 | 0.0000 | 0.0000 |
| 16 | 0.1853 | 0.3294 | 0.2745 | 0.1423 | 0.0514 | 0.0137 | 0.0028 | 0.0004 | 0.0001 | 0.0000 |
| 17 | 0.1668 | 0.3150 | 0.2800 | 0.1556 | 0.0605 | 0.0175 | 0.0039 | 0.0007 | 0.0001 | 0.0000 |
| 18 | 0.1501 | 0.3002 | 0.2835 | 0.1680 | 0.0700 | 0.0218 | 0.0052 | 0.0010 | 0.0002 | 0.0000 |
| 19 | 0.1351 | 0.2852 | 0.2852 | 0.1796 | 0.0798 | 0.0266 | 0.0069 | 0.0014 | 0.0002 | 0.0000 |
| 20 | 0.1216 | 0.2702 | 0.2852 | 0.1901 | 0.0898 | 0.0319 | 0.0089 | 0.0020 | 0.0004 | 0.0001 |
| 21 | 0.1094 | 0.2553 | 0.2837 | 0.1996 | 0.0998 | 0.0377 | 0.0112 | 0.0027 | 0.0005 | 0.0001 |
| 22 | 0.0985 | 0.2407 | 0.2808 | 0.2080 | 0.1098 | 0.0439 | 0.0138 | 0.0035 | 0.0007 | 0.0001 |
| 23 | 0.0886 | 0.2265 | 0.2768 | 0.2153 | 0.1196 | 0.0505 | 0.0168 | 0.0045 | 0.0010 | 0.0002 |
| 24 | 0.0798 | 0.2127 | 0.2718 | 0.2215 | 0.1292 | 0.0574 | 0.0202 | 0.0058 | 0.0014 | 0.0003 |
| 25 | 0.0718 | 0.1994 | 0.2659 | 0.2265 | 0.1384 | 0.0646 | 0.0239 | 0.0072 | 0.0018 | 0.0004 |
| 26 | 0.0646 | 0.1867 | 0.2592 | 0.2304 | 0.1472 | 0.0720 | 0.0280 | 0.0089 | 0.0023 | 0.0005 |
| 27 | 0.0581 | 0.1744 | 0.2520 | 0.2333 | 0.1555 | 0.0795 | 0.0324 | 0.0108 | 0.0030 | 0.0007 |
| 28 | 0.0523 | 0.1628 | 0.2442 | 0.2352 | 0.1633 | 0.0871 | 0.0371 | 0.0130 | 0.0038 | 0.0009 |
| 29 | 0.0471 | 0.1518 | 0.2361 | 0.2361 | 0.1705 | 0.0947 | 0.0421 | 0.0154 | 0.0047 | 0.0012 |
| 30 | 0.0424 | 0.1413 | 0.2277 | 0.2361 | 0.1771 | 0.1023 | 0.0474 | 0.0180 | 0.0058 | 0.0016 |
| 32 | 0.0343 | 0.1221 | 0.2103 | 0.2336 | 0.1882 | 0.1171 | 0.0585 | 0.0242 | 0.0084 | 0.0025 |
| 34 | 0.0278 | 0.1051 | 0.1926 | 0.2283 | 0.1966 | 0.1311 | 0.0704 | 0.0313 | 0.0117 | 0.0038 |
| 36 | 0.0225 | 0.0901 | 0.1752 | 0.2206 | 0.2023 | 0.1438 | 0.0826 | 0.0393 | 0.0158 | 0.0055 |
| 38 | 0.0182 | 0.0770 | 0.1584 | 0.2112 | 0.2053 | 0.1551 | 0.0948 | 0.0481 | 0.0207 | 0.0077 |
| 40 | 0.0148 | 0.0657 | 0.1423 | 0.2003 | 0.2059 | 0.1647 | 0.1068 | 0.0576 | 0.0264 | 0.0104 |

Table T-8: Selected binomial probabilities (Continued)

| n | y = number of attributes in a binomial sample of size n with $\pi = 0.25$ | | | | | | | | | |
|----|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.7500 | 0.2500 | | | | | | | | |
| 2 | 0.5625 | 0.3750 | 0.0625 | | | | | | | |
| 3 | 0.4219 | 0.4219 | 0.1406 | 0.0156 | | | | | | |
| 4 | 0.3164 | 0.4219 | 0.2109 | 0.0469 | 0.0039 | | | | | |
| 5 | 0.2373 | 0.3955 | 0.2637 | 0.0879 | 0.0146 | 0.0010 | | | | |
| 6 | 0.1780 | 0.3560 | 0.2966 | 0.1318 | 0.0330 | 0.0044 | 0.0002 | | | |
| 7 | 0.1335 | 0.3115 | 0.3115 | 0.1730 | 0.0577 | 0.0115 | 0.0013 | 0.0001 | | |
| 8 | 0.1001 | 0.2670 | 0.3115 | 0.2076 | 0.0865 | 0.0231 | 0.0038 | 0.0004 | 0.0000 | |
| 9 | 0.0751 | 0.2253 | 0.3003 | 0.2336 | 0.1168 | 0.0389 | 0.0087 | 0.0012 | 0.0001 | 0.0000 |
| 10 | 0.0563 | 0.1877 | 0.2816 | 0.2503 | 0.1460 | 0.0584 | 0.0162 | 0.0031 | 0.0004 | 0.0000 |
| 11 | 0.0422 | 0.1549 | 0.2581 | 0.2581 | 0.1721 | 0.0803 | 0.0268 | 0.0064 | 0.0011 | 0.0001 |
| 12 | 0.0317 | 0.1267 | 0.2323 | 0.2581 | 0.1936 | 0.1032 | 0.0401 | 0.0115 | 0.0024 | 0.0004 |
| 13 | 0.0238 | 0.1029 | 0.2059 | 0.2517 | 0.2097 | 0.1258 | 0.0559 | 0.0186 | 0.0047 | 0.0009 |
| 14 | 0.0178 | 0.0832 | 0.1802 | 0.2402 | 0.2202 | 0.1468 | 0.0734 | 0.0280 | 0.0082 | 0.0018 |
| 15 | 0.0134 | 0.0668 | 0.1559 | 0.2252 | 0.2252 | 0.1651 | 0.0917 | 0.0393 | 0.0131 | 0.0034 |
| 16 | 0.0100 | 0.0535 | 0.1336 | 0.2079 | 0.2252 | 0.1802 | 0.1101 | 0.0524 | 0.0197 | 0.0058 |
| 17 | 0.0075 | 0.0426 | 0.1136 | 0.1893 | 0.2209 | 0.1914 | 0.1276 | 0.0668 | 0.0279 | 0.0093 |
| 18 | 0.0056 | 0.0338 | 0.0958 | 0.1704 | 0.2130 | 0.1988 | 0.1436 | 0.0820 | 0.0376 | 0.0139 |
| 19 | 0.0042 | 0.0268 | 0.0803 | 0.1517 | 0.2023 | 0.2023 | 0.1574 | 0.0974 | 0.0487 | 0.0198 |
| 20 | 0.0032 | 0.0211 | 0.0669 | 0.1339 | 0.1897 | 0.2023 | 0.1686 | 0.1124 | 0.0609 | 0.0271 |
| 21 | 0.0024 | 0.0166 | 0.0555 | 0.1172 | 0.1757 | 0.1992 | 0.1770 | 0.1265 | 0.0738 | 0.0355 |
| 22 | 0.0018 | 0.0131 | 0.0458 | 0.1017 | 0.1611 | 0.1933 | 0.1826 | 0.1391 | 0.0869 | 0.0451 |
| 23 | 0.0013 | 0.0103 | 0.0376 | 0.0878 | 0.1463 | 0.1853 | 0.1853 | 0.1500 | 0.1000 | 0.0555 |
| 24 | 0.0010 | 0.0080 | 0.0308 | 0.0752 | 0.1316 | 0.1755 | 0.1853 | 0.1588 | 0.1125 | 0.0667 |
| 25 | 0.0008 | 0.0063 | 0.0251 | 0.0641 | 0.1175 | 0.1645 | 0.1828 | 0.1654 | 0.1241 | 0.0781 |
| 26 | 0.0006 | 0.0049 | 0.0204 | 0.0544 | 0.1042 | 0.1528 | 0.1782 | 0.1698 | 0.1344 | 0.0896 |
| 27 | 0.0004 | 0.0038 | 0.0165 | 0.0459 | 0.0917 | 0.1406 | 0.1719 | 0.1719 | 0.1432 | 0.1008 |
| 28 | 0.0003 | 0.0030 | 0.0133 | 0.0385 | 0.0803 | 0.1284 | 0.1641 | 0.1719 | 0.1504 | 0.1114 |
| 29 | 0.0002 | 0.0023 | 0.0107 | 0.0322 | 0.0698 | 0.1164 | 0.1552 | 0.1699 | 0.1558 | 0.1212 |
| 30 | 0.0002 | 0.0018 | 0.0086 | 0.0269 | 0.0604 | 0.1047 | 0.1455 | 0.1662 | 0.1593 | 0.1298 |
| 32 | 0.0001 | 0.0011 | 0.0055 | 0.0185 | 0.0446 | 0.0832 | 0.1249 | 0.1546 | 0.1610 | 0.1431 |
| 34 | 0.0001 | 0.0006 | 0.0035 | 0.0125 | 0.0324 | 0.0647 | 0.1042 | 0.1390 | 0.1564 | 0.1506 |
| 36 | 0.0000 | 0.0004 | 0.0022 | 0.0084 | 0.0231 | 0.0493 | 0.0849 | 0.1213 | 0.1466 | 0.1520 |
| 38 | 0.0000 | 0.0002 | 0.0014 | 0.0056 | 0.0163 | 0.0369 | 0.0677 | 0.1032 | 0.1333 | 0.1481 |
| 40 | 0.0000 | 0.0001 | 0.0009 | 0.0037 | 0.0113 | 0.0272 | 0.0530 | 0.0857 | 0.1179 | 0.1397 |

Table T-8: Selected binomial probabilities (Continued)

| <i>n</i> | <i>y</i> = number of attributes in a binomial sample of size <i>n</i> with $\pi = 0.50$ | | | | | | | | | |
|----------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.5000 | 0.5000 | | | | | | | | |
| 2 | 0.2500 | 0.5000 | 0.2500 | | | | | | | |
| 3 | 0.1250 | 0.3750 | 0.3750 | 0.1250 | | | | | | |
| 4 | 0.0625 | 0.2500 | 0.3750 | 0.2500 | 0.0625 | | | | | |
| 5 | 0.0313 | 0.1563 | 0.3125 | 0.3125 | 0.1563 | 0.0313 | | | | |
| 6 | 0.0156 | 0.0938 | 0.2344 | 0.3125 | 0.2344 | 0.0938 | 0.0156 | | | |
| 7 | 0.0078 | 0.0547 | 0.1641 | 0.2734 | 0.2734 | 0.1641 | 0.0547 | 0.0078 | | |
| 8 | 0.0039 | 0.0313 | 0.1094 | 0.2188 | 0.2734 | 0.2188 | 0.1094 | 0.0313 | 0.0039 | |
| 9 | 0.0020 | 0.0176 | 0.0703 | 0.1641 | 0.2461 | 0.2461 | 0.1641 | 0.0703 | 0.0176 | 0.0020 |
| 10 | 0.0010 | 0.0098 | 0.0439 | 0.1172 | 0.2051 | 0.2461 | 0.2051 | 0.1172 | 0.0439 | 0.0098 |
| 11 | 0.0005 | 0.0054 | 0.0269 | 0.0806 | 0.1611 | 0.2256 | 0.2256 | 0.1611 | 0.0806 | 0.0269 |
| 12 | 0.0002 | 0.0029 | 0.0161 | 0.0537 | 0.1208 | 0.1934 | 0.2256 | 0.1934 | 0.1208 | 0.0537 |
| 13 | 0.0001 | 0.0016 | 0.0095 | 0.0349 | 0.0873 | 0.1571 | 0.2095 | 0.2095 | 0.1571 | 0.0873 |
| 14 | 0.0001 | 0.0009 | 0.0056 | 0.0222 | 0.0611 | 0.1222 | 0.1833 | 0.2095 | 0.1833 | 0.1222 |
| 15 | 0.0000 | 0.0005 | 0.0032 | 0.0139 | 0.0417 | 0.0916 | 0.1527 | 0.1964 | 0.1964 | 0.1527 |
| 16 | 0.0000 | 0.0002 | 0.0018 | 0.0085 | 0.0278 | 0.0667 | 0.1222 | 0.1746 | 0.1964 | 0.1746 |
| 17 | 0.0000 | 0.0001 | 0.0010 | 0.0052 | 0.0182 | 0.0472 | 0.0944 | 0.1484 | 0.1855 | 0.1855 |
| 18 | 0.0000 | 0.0001 | 0.0006 | 0.0031 | 0.0117 | 0.0327 | 0.0708 | 0.1214 | 0.1669 | 0.1855 |
| 19 | 0.0000 | 0.0000 | 0.0003 | 0.0018 | 0.0074 | 0.0222 | 0.0518 | 0.0961 | 0.1442 | 0.1762 |
| 20 | 0.0000 | 0.0000 | 0.0002 | 0.0011 | 0.0046 | 0.0148 | 0.0370 | 0.0739 | 0.1201 | 0.1602 |
| 21 | 0.0000 | 0.0000 | 0.0001 | 0.0006 | 0.0029 | 0.0097 | 0.0259 | 0.0554 | 0.0970 | 0.1402 |
| 22 | 0.0000 | 0.0000 | 0.0001 | 0.0004 | 0.0017 | 0.0063 | 0.0178 | 0.0407 | 0.0762 | 0.1186 |
| 23 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0011 | 0.0040 | 0.0120 | 0.0292 | 0.0584 | 0.0974 |
| 24 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0006 | 0.0025 | 0.0080 | 0.0206 | 0.0438 | 0.0779 |
| 25 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0004 | 0.0016 | 0.0053 | 0.0143 | 0.0322 | 0.0609 |
| 26 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0010 | 0.0034 | 0.0098 | 0.0233 | 0.0466 |
| 27 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0006 | 0.0022 | 0.0066 | 0.0165 | 0.0349 |
| 28 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0004 | 0.0014 | 0.0044 | 0.0116 | 0.0257 |
| 29 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0009 | 0.0029 | 0.0080 | 0.0187 |
| 30 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0006 | 0.0019 | 0.0055 | 0.0133 |
| 32 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0008 | 0.0024 | 0.0065 |
| 34 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0011 | 0.0031 |
| 36 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0004 | 0.0014 |
| 38 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0006 |
| 40 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 |

Table T-9: Selected Poisson probabilities (continued)

| y | λ | | | | | | | | | |
|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 5.1 | 5.2 | 5.3 | 5.4 | 5.5 | 5.6 | 5.7 | 5.8 | 5.9 | 6.0 |
| 0 | 0.0061 | 0.0055 | 0.0050 | 0.0045 | 0.0041 | 0.0037 | 0.0033 | 0.0030 | 0.0027 | 0.0025 |
| 1 | 0.0311 | 0.0287 | 0.0265 | 0.0244 | 0.0225 | 0.0207 | 0.0191 | 0.0176 | 0.0162 | 0.0149 |
| 2 | 0.0793 | 0.0746 | 0.0701 | 0.0659 | 0.0618 | 0.0580 | 0.0544 | 0.0509 | 0.0477 | 0.0446 |
| 3 | 0.1348 | 0.1293 | 0.1239 | 0.1185 | 0.1133 | 0.1082 | 0.1033 | 0.0985 | 0.0938 | 0.0892 |
| 4 | 0.1719 | 0.1681 | 0.1641 | 0.1600 | 0.1558 | 0.1515 | 0.1472 | 0.1428 | 0.1383 | 0.1339 |
| 5 | 0.1753 | 0.1748 | 0.1740 | 0.1728 | 0.1714 | 0.1697 | 0.1678 | 0.1656 | 0.1632 | 0.1606 |
| 6 | 0.1490 | 0.1515 | 0.1537 | 0.1555 | 0.1571 | 0.1584 | 0.1594 | 0.1601 | 0.1605 | 0.1606 |
| 7 | 0.1086 | 0.1125 | 0.1163 | 0.1200 | 0.1234 | 0.1267 | 0.1298 | 0.1326 | 0.1353 | 0.1377 |
| 8 | 0.0692 | 0.0731 | 0.0771 | 0.0810 | 0.0849 | 0.0887 | 0.0925 | 0.0962 | 0.0998 | 0.1033 |
| 9 | 0.0392 | 0.0423 | 0.0454 | 0.0486 | 0.0519 | 0.0552 | 0.0586 | 0.0620 | 0.0654 | 0.0688 |
| 10 | 0.0200 | 0.0220 | 0.0241 | 0.0262 | 0.0285 | 0.0309 | 0.0334 | 0.0359 | 0.0386 | 0.0413 |
| 11 | 0.0093 | 0.0104 | 0.0116 | 0.0129 | 0.0143 | 0.0157 | 0.0173 | 0.0190 | 0.0207 | 0.0225 |
| 12 | 0.0039 | 0.0045 | 0.0051 | 0.0058 | 0.0065 | 0.0073 | 0.0082 | 0.0092 | 0.0102 | 0.0113 |
| 13 | 0.0015 | 0.0018 | 0.0021 | 0.0024 | 0.0028 | 0.0032 | 0.0036 | 0.0041 | 0.0046 | 0.0052 |
| 14 | 0.0006 | 0.0007 | 0.0008 | 0.0009 | 0.0011 | 0.0013 | 0.0015 | 0.0017 | 0.0019 | 0.0022 |
| 15 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0004 | 0.0005 | 0.0006 | 0.0007 | 0.0008 | 0.0009 |
| 16 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0002 | 0.0003 | 0.0003 |
| 17 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |

Table T-9: Selected Poisson probabilities (continued)

| y | λ | | | | | | | | | |
|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 6.1 | 6.2 | 6.3 | 6.4 | 6.5 | 6.6 | 6.7 | 6.8 | 6.9 | 7.0 |
| 0 | 0.0022 | 0.0020 | 0.0018 | 0.0017 | 0.0015 | 0.0014 | 0.0012 | 0.0011 | 0.0010 | 0.0009 |
| 1 | 0.0137 | 0.0126 | 0.0116 | 0.0106 | 0.0098 | 0.0090 | 0.0082 | 0.0076 | 0.0070 | 0.0064 |
| 2 | 0.0417 | 0.0390 | 0.0364 | 0.0340 | 0.0318 | 0.0296 | 0.0276 | 0.0258 | 0.0240 | 0.0223 |
| 3 | 0.0848 | 0.0806 | 0.0765 | 0.0726 | 0.0688 | 0.0652 | 0.0617 | 0.0584 | 0.0552 | 0.0521 |
| 4 | 0.1294 | 0.1249 | 0.1205 | 0.1162 | 0.1118 | 0.1076 | 0.1034 | 0.0992 | 0.0952 | 0.0912 |
| 5 | 0.1579 | 0.1549 | 0.1519 | 0.1487 | 0.1454 | 0.1420 | 0.1385 | 0.1349 | 0.1314 | 0.1277 |
| 6 | 0.1605 | 0.1601 | 0.1595 | 0.1586 | 0.1575 | 0.1562 | 0.1546 | 0.1529 | 0.1511 | 0.1490 |
| 7 | 0.1399 | 0.1418 | 0.1435 | 0.1450 | 0.1462 | 0.1472 | 0.1480 | 0.1486 | 0.1489 | 0.1490 |
| 8 | 0.1066 | 0.1099 | 0.1130 | 0.1160 | 0.1188 | 0.1215 | 0.1240 | 0.1263 | 0.1284 | 0.1304 |
| 9 | 0.0723 | 0.0757 | 0.0791 | 0.0825 | 0.0858 | 0.0891 | 0.0923 | 0.0954 | 0.0985 | 0.1014 |
| 10 | 0.0441 | 0.0469 | 0.0498 | 0.0528 | 0.0558 | 0.0588 | 0.0618 | 0.0649 | 0.0679 | 0.0710 |
| 11 | 0.0244 | 0.0265 | 0.0285 | 0.0307 | 0.0330 | 0.0353 | 0.0377 | 0.0401 | 0.0426 | 0.0452 |
| 12 | 0.0124 | 0.0137 | 0.0150 | 0.0164 | 0.0179 | 0.0194 | 0.0210 | 0.0227 | 0.0245 | 0.0263 |
| 13 | 0.0058 | 0.0065 | 0.0073 | 0.0081 | 0.0089 | 0.0099 | 0.0108 | 0.0119 | 0.0130 | 0.0142 |
| 14 | 0.0025 | 0.0029 | 0.0033 | 0.0037 | 0.0041 | 0.0046 | 0.0052 | 0.0058 | 0.0064 | 0.0071 |
| 15 | 0.0010 | 0.0012 | 0.0014 | 0.0016 | 0.0018 | 0.0020 | 0.0023 | 0.0026 | 0.0029 | 0.0033 |
| 16 | 0.0004 | 0.0005 | 0.0005 | 0.0006 | 0.0007 | 0.0008 | 0.0010 | 0.0011 | 0.0013 | 0.0014 |
| 17 | 0.0001 | 0.0002 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0004 | 0.0004 | 0.0005 | 0.0006 |
| 18 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0002 |
| 19 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |

Table T-9: Selected Poisson probabilities (continued)

| y | λ | | | | | | | | | |
|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 7.1 | 7.2 | 7.3 | 7.4 | 7.5 | 7.6 | 7.7 | 7.8 | 7.9 | 8.0 |
| 0 | 0.0008 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0003 |
| 1 | 0.0059 | 0.0054 | 0.0049 | 0.0045 | 0.0041 | 0.0038 | 0.0035 | 0.0032 | 0.0029 | 0.0027 |
| 2 | 0.0208 | 0.0194 | 0.0180 | 0.0167 | 0.0156 | 0.0145 | 0.0134 | 0.0125 | 0.0116 | 0.0107 |
| 3 | 0.0492 | 0.0464 | 0.0438 | 0.0413 | 0.0389 | 0.0366 | 0.0345 | 0.0324 | 0.0305 | 0.0286 |
| 4 | 0.0874 | 0.0836 | 0.0799 | 0.0764 | 0.0729 | 0.0696 | 0.0663 | 0.0632 | 0.0602 | 0.0573 |
| 5 | 0.1241 | 0.1204 | 0.1167 | 0.1130 | 0.1094 | 0.1057 | 0.1021 | 0.0986 | 0.0951 | 0.0916 |
| 6 | 0.1468 | 0.1445 | 0.1420 | 0.1394 | 0.1367 | 0.1339 | 0.1311 | 0.1282 | 0.1252 | 0.1221 |
| 7 | 0.1489 | 0.1486 | 0.1481 | 0.1474 | 0.1465 | 0.1454 | 0.1442 | 0.1428 | 0.1413 | 0.1396 |
| 8 | 0.1321 | 0.1337 | 0.1351 | 0.1363 | 0.1373 | 0.1381 | 0.1388 | 0.1392 | 0.1395 | 0.1396 |
| 9 | 0.1042 | 0.1070 | 0.1096 | 0.1121 | 0.1144 | 0.1167 | 0.1187 | 0.1207 | 0.1224 | 0.1241 |
| 10 | 0.0740 | 0.0770 | 0.0800 | 0.0829 | 0.0858 | 0.0887 | 0.0914 | 0.0941 | 0.0967 | 0.0993 |
| 11 | 0.0478 | 0.0504 | 0.0531 | 0.0558 | 0.0585 | 0.0613 | 0.0640 | 0.0667 | 0.0695 | 0.0722 |
| 12 | 0.0283 | 0.0303 | 0.0323 | 0.0344 | 0.0366 | 0.0388 | 0.0411 | 0.0434 | 0.0457 | 0.0481 |
| 13 | 0.0154 | 0.0168 | 0.0181 | 0.0196 | 0.0211 | 0.0227 | 0.0243 | 0.0260 | 0.0278 | 0.0296 |
| 14 | 0.0078 | 0.0086 | 0.0095 | 0.0104 | 0.0113 | 0.0123 | 0.0134 | 0.0145 | 0.0157 | 0.0169 |
| 15 | 0.0037 | 0.0041 | 0.0046 | 0.0051 | 0.0057 | 0.0062 | 0.0069 | 0.0075 | 0.0083 | 0.0090 |
| 16 | 0.0016 | 0.0019 | 0.0021 | 0.0024 | 0.0026 | 0.0030 | 0.0033 | 0.0037 | 0.0041 | 0.0045 |
| 17 | 0.0007 | 0.0008 | 0.0009 | 0.0010 | 0.0012 | 0.0013 | 0.0015 | 0.0017 | 0.0019 | 0.0021 |
| 18 | 0.0003 | 0.0003 | 0.0004 | 0.0004 | 0.0005 | 0.0006 | 0.0006 | 0.0007 | 0.0008 | 0.0009 |
| 19 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0003 | 0.0004 |
| 20 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 |
| 21 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 |

Table T-9 Selected Poisson probabilities (continued)

| y | λ | | | | | | | | | |
|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 9.1 | 9.2 | 9.3 | 9.4 | 9.5 | 9.6 | 9.7 | 9.8 | 9.9 | 10.0 |
| 0 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0000 |
| 1 | 0.0010 | 0.0009 | 0.0009 | 0.0008 | 0.0007 | 0.0007 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| 2 | 0.0046 | 0.0043 | 0.0040 | 0.0037 | 0.0034 | 0.0031 | 0.0029 | 0.0027 | 0.0025 | 0.0023 |
| 3 | 0.0140 | 0.0131 | 0.0123 | 0.0115 | 0.0107 | 0.0100 | 0.0093 | 0.0087 | 0.0081 | 0.0076 |
| 4 | 0.0319 | 0.0302 | 0.0285 | 0.0269 | 0.0254 | 0.0240 | 0.0226 | 0.0213 | 0.0201 | 0.0189 |
| 5 | 0.0581 | 0.0555 | 0.0530 | 0.0506 | 0.0483 | 0.0460 | 0.0439 | 0.0418 | 0.0398 | 0.0378 |
| 6 | 0.0881 | 0.0851 | 0.0822 | 0.0793 | 0.0764 | 0.0736 | 0.0709 | 0.0682 | 0.0656 | 0.0631 |
| 7 | 0.1145 | 0.1118 | 0.1091 | 0.1064 | 0.1037 | 0.1010 | 0.0982 | 0.0955 | 0.0928 | 0.0901 |
| 8 | 0.1302 | 0.1286 | 0.1269 | 0.1251 | 0.1232 | 0.1212 | 0.1191 | 0.1170 | 0.1148 | 0.1126 |
| 9 | 0.1317 | 0.1315 | 0.1311 | 0.1306 | 0.1300 | 0.1293 | 0.1284 | 0.1274 | 0.1263 | 0.1251 |
| 10 | 0.1198 | 0.1210 | 0.1219 | 0.1228 | 0.1235 | 0.1241 | 0.1245 | 0.1249 | 0.1250 | 0.1251 |
| 11 | 0.0991 | 0.1012 | 0.1031 | 0.1049 | 0.1067 | 0.1083 | 0.1098 | 0.1112 | 0.1125 | 0.1137 |
| 12 | 0.0752 | 0.0776 | 0.0799 | 0.0822 | 0.0844 | 0.0866 | 0.0888 | 0.0908 | 0.0928 | 0.0948 |
| 13 | 0.0526 | 0.0549 | 0.0572 | 0.0594 | 0.0617 | 0.0640 | 0.0662 | 0.0685 | 0.0707 | 0.0729 |
| 14 | 0.0342 | 0.0361 | 0.0380 | 0.0399 | 0.0419 | 0.0439 | 0.0459 | 0.0479 | 0.0500 | 0.0521 |
| 15 | 0.0208 | 0.0221 | 0.0235 | 0.0250 | 0.0265 | 0.0281 | 0.0297 | 0.0313 | 0.0330 | 0.0347 |
| 16 | 0.0118 | 0.0127 | 0.0137 | 0.0147 | 0.0157 | 0.0168 | 0.0180 | 0.0192 | 0.0204 | 0.0217 |
| 17 | 0.0063 | 0.0069 | 0.0075 | 0.0081 | 0.0088 | 0.0095 | 0.0103 | 0.0111 | 0.0119 | 0.0128 |
| 18 | 0.0032 | 0.0035 | 0.0039 | 0.0042 | 0.0046 | 0.0051 | 0.0055 | 0.0060 | 0.0065 | 0.0071 |
| 19 | 0.0015 | 0.0017 | 0.0019 | 0.0021 | 0.0023 | 0.0026 | 0.0028 | 0.0031 | 0.0034 | 0.0037 |
| 20 | 0.0007 | 0.0008 | 0.0009 | 0.0010 | 0.0011 | 0.0012 | 0.0014 | 0.0015 | 0.0017 | 0.0019 |
| 21 | 0.0003 | 0.0003 | 0.0004 | 0.0004 | 0.0005 | 0.0006 | 0.0006 | 0.0007 | 0.0008 | 0.0009 |
| 22 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0004 | 0.0004 |
| 23 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 |
| 24 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 |

Table T-9 Selected Poisson probabilities (continued)

| y | λ | | | | | | | | | |
|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 11.0 | 12.0 | 13.0 | 14.0 | 15.0 | 16.0 | 17.0 | 18.0 | 19.0 | 20.0 |
| 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 0.0010 | 0.0004 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3 | 0.0037 | 0.0018 | 0.0008 | 0.0004 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | 0.0102 | 0.0053 | 0.0027 | 0.0013 | 0.0006 | 0.0003 | 0.0001 | 0.0001 | 0.0000 | 0.0000 |
| 5 | 0.0224 | 0.0127 | 0.0070 | 0.0037 | 0.0019 | 0.0010 | 0.0005 | 0.0002 | 0.0001 | 0.0001 |
| 6 | 0.0411 | 0.0255 | 0.0152 | 0.0087 | 0.0048 | 0.0026 | 0.0014 | 0.0007 | 0.0004 | 0.0002 |
| 7 | 0.0646 | 0.0437 | 0.0281 | 0.0174 | 0.0104 | 0.0060 | 0.0034 | 0.0019 | 0.0010 | 0.0005 |
| 8 | 0.0888 | 0.0655 | 0.0457 | 0.0304 | 0.0194 | 0.0120 | 0.0072 | 0.0042 | 0.0024 | 0.0013 |
| 9 | 0.1085 | 0.0874 | 0.0661 | 0.0473 | 0.0324 | 0.0213 | 0.0135 | 0.0083 | 0.0050 | 0.0029 |
| 10 | 0.1194 | 0.1048 | 0.0859 | 0.0663 | 0.0486 | 0.0341 | 0.0230 | 0.0150 | 0.0095 | 0.0058 |
| 11 | 0.1194 | 0.1144 | 0.1015 | 0.0844 | 0.0663 | 0.0496 | 0.0355 | 0.0245 | 0.0164 | 0.0106 |
| 12 | 0.1094 | 0.1144 | 0.1099 | 0.0984 | 0.0829 | 0.0661 | 0.0504 | 0.0368 | 0.0259 | 0.0176 |
| 13 | 0.0926 | 0.1056 | 0.1099 | 0.1060 | 0.0956 | 0.0814 | 0.0658 | 0.0509 | 0.0378 | 0.0271 |
| 14 | 0.0728 | 0.0905 | 0.1021 | 0.1060 | 0.1024 | 0.0930 | 0.0800 | 0.0655 | 0.0514 | 0.0387 |
| 15 | 0.0534 | 0.0724 | 0.0885 | 0.0989 | 0.1024 | 0.0992 | 0.0906 | 0.0786 | 0.0650 | 0.0516 |
| 16 | 0.0367 | 0.0543 | 0.0719 | 0.0866 | 0.0960 | 0.0992 | 0.0963 | 0.0884 | 0.0772 | 0.0646 |
| 17 | 0.0237 | 0.0383 | 0.0550 | 0.0713 | 0.0847 | 0.0934 | 0.0963 | 0.0936 | 0.0863 | 0.0760 |
| 18 | 0.0145 | 0.0255 | 0.0397 | 0.0554 | 0.0706 | 0.0830 | 0.0909 | 0.0936 | 0.0911 | 0.0844 |
| 19 | 0.0084 | 0.0161 | 0.0272 | 0.0409 | 0.0557 | 0.0699 | 0.0814 | 0.0887 | 0.0911 | 0.0888 |
| 20 | 0.0046 | 0.0097 | 0.0177 | 0.0286 | 0.0418 | 0.0559 | 0.0692 | 0.0798 | 0.0866 | 0.0888 |
| 21 | 0.0024 | 0.0055 | 0.0109 | 0.0191 | 0.0299 | 0.0426 | 0.0560 | 0.0684 | 0.0783 | 0.0846 |
| 22 | 0.0012 | 0.0030 | 0.0065 | 0.0121 | 0.0204 | 0.0310 | 0.0433 | 0.0560 | 0.0676 | 0.0769 |
| 23 | 0.0006 | 0.0016 | 0.0037 | 0.0074 | 0.0133 | 0.0216 | 0.0320 | 0.0438 | 0.0559 | 0.0669 |
| 24 | 0.0003 | 0.0008 | 0.0020 | 0.0043 | 0.0083 | 0.0144 | 0.0226 | 0.0328 | 0.0442 | 0.0557 |
| 25 | 0.0001 | 0.0004 | 0.0010 | 0.0024 | 0.0050 | 0.0092 | 0.0154 | 0.0237 | 0.0336 | 0.0446 |
| 26 | 0.0000 | 0.0002 | 0.0005 | 0.0013 | 0.0029 | 0.0057 | 0.0101 | 0.0164 | 0.0246 | 0.0343 |
| 27 | 0.0000 | 0.0001 | 0.0002 | 0.0007 | 0.0016 | 0.0034 | 0.0063 | 0.0109 | 0.0173 | 0.0254 |
| 28 | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0009 | 0.0019 | 0.0038 | 0.0070 | 0.0117 | 0.0181 |
| 29 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0004 | 0.0011 | 0.0023 | 0.0044 | 0.0077 | 0.0125 |
| 30 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0006 | 0.0013 | 0.0026 | 0.0049 | 0.0083 |
| 31 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0007 | 0.0015 | 0.0030 | 0.0054 |
| 32 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0004 | 0.0009 | 0.0018 | 0.0034 |
| 33 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0005 | 0.0010 | 0.0020 |
| 34 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0006 | 0.0012 |
| 35 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0007 |

Table T-10: Confidence limits for the Poisson parameter λ
(see Section 23.5)

| 1 - α | 0.998 | | 0.99 | | 0.95 | | 0.90 | |
|--------------------------------|--------------|-------------|--------------|-------------|--------------|-------------|-------------|-------------|
| $\alpha/2$ | 0.001 | | 0.005 | | 0.025 | | 0.05 | |
| y | λ_L | λ_U | λ_L | λ_U | λ_L | λ_U | λ_L | λ_U |
| 0 | 0 | 6.9075 | 0 | 5.2983 | 0 | 3.6889 | 0 | 2.9957 |
| 1 | 0.0011 | 9.2331 | 0.0051 | 7.4301 | 0.0251 | 5.5721 | 0.0511 | 4.7441 |
| 2 | 0.0454 | 11.2287 | 0.1035 | 9.2738 | 0.2422 | 7.2247 | 0.3554 | 6.2958 |
| 3 | 0.191 | 13.062 | 0.338 | 10.997 | 0.619 | 8.767 | 0.818 | 7.754 |
| 4 | 0.429 | 14.794 | 0.672 | 12.594 | 1.09 | 10.242 | 1.366 | 9.154 |
| 5 | 0.739 | 16.45 | 1.08 | 14.15 | 1.62 | 11.67 | 1.97 | 10.51 |
| 6 | 1.11 | 18.06 | 1.54 | 15.66 | 2.2 | 13.06 | 2.61 | 11.84 |
| 7 | 1.52 | 19.63 | 2.04 | 17.13 | 2.81 | 14.42 | 3.29 | 13.15 |
| 8 | 1.97 | 21.16 | 2.57 | 18.58 | 3.45 | 15.76 | 3.98 | 14.43 |
| 9 | 2.45 | 22.66 | 3.13 | 20 | 4.12 | 17.08 | 4.7 | 15.71 |
| 10 | 2.96 | 24.13 | 3.72 | 21.4 | 4.8 | 18.39 | 5.43 | 16.96 |
| 11 | 3.49 | 25.59 | 4.32 | 22.78 | 5.49 | 19.68 | 6.17 | 18.21 |
| 12 | 4.04 | 27.03 | 4.94 | 24.14 | 6.2 | 20.96 | 6.92 | 19.44 |
| 13 | 4.61 | 28.45 | 5.58 | 25.5 | 6.92 | 22.23 | 7.69 | 20.67 |
| 14 | 5.2 | 29.85 | 6.23 | 26.84 | 7.65 | 23.49 | 8.46 | 21.89 |
| 15 | 5.79 | 31.24 | 6.89 | 28.16 | 8.4 | 24.74 | 9.25 | 23.1 |
| 16 | 6.41 | 32.62 | 7.57 | 29.48 | 9.15 | 25.98 | 10.04 | 24.3 |
| 17 | 7.03 | 33.99 | 8.25 | 30.79 | 9.9 | 27.22 | 10.83 | 25.5 |
| 18 | 7.66 | 35.35 | 8.94 | 32.09 | 10.67 | 28.45 | 11.63 | 26.69 |
| 19 | 8.31 | 36.7 | 9.64 | 33.38 | 11.44 | 29.67 | 12.44 | 27.88 |
| 20 | 8.96 | 38.04 | 10.35 | 34.67 | 12.22 | 30.89 | 13.25 | 29.06 |
| 21 | 9.62 | 39.37 | 11.07 | 35.95 | 13 | 32.1 | 14.07 | 30.24 |
| 22 | 10.29 | 40.7 | 11.79 | 37.22 | 13.79 | 33.31 | 14.89 | 31.41 |
| 23 | 10.96 | 42.02 | 12.52 | 38.48 | 14.58 | 34.51 | 15.72 | 32.59 |
| 24 | 11.65 | 43.33 | 13.25 | 39.74 | 15.38 | 35.71 | 16.55 | 33.75 |
| 25 | 12.34 | 44.64 | 14 | 41 | 16.18 | 36.9 | 17.38 | 34.92 |
| 26 | 13.03 | 45.94 | 14.74 | 42.25 | 16.98 | 38.1 | 18.22 | 36.08 |
| 27 | 13.73 | 47.23 | 15.49 | 43.5 | 17.79 | 39.28 | 19.06 | 37.23 |
| 28 | 14.44 | 48.52 | 16.25 | 44.74 | 18.61 | 40.47 | 19.9 | 38.39 |
| 29 | 15.15 | 49.8 | 17 | 45.98 | 19.42 | 41.65 | 20.75 | 39.54 |
| 30 | 15.87 | 51.08 | 17.77 | 47.21 | 20.24 | 42.83 | 21.59 | 40.69 |
| 35 | 19.52 | 57.42 | 21.64 | 53.32 | 24.38 | 48.68 | 25.87 | 46.4 |
| 40 | 23.26 | 63.66 | 25.59 | 59.36 | 28.58 | 54.47 | 30.2 | 52.07 |
| 45 | 27.08 | 69.83 | 29.6 | 65.34 | 32.82 | 60.21 | 34.56 | 57.69 |
| 50 | 30.96 | 75.94 | 33.66 | 71.27 | 37.11 | 65.92 | 38.96 | 63.29 |

Table T-11a: Two-sided tolerance limit factors for a normal distribution
(see Section 9.12)

| n | $\gamma = 0.90$ | | | $\gamma = 0.95$ | | | $\gamma = 0.99$ | | |
|----------|-----------------|--------------|--------------|-----------------|--------------|--------------|-----------------|--------------|--------------|
| | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ |
| 2 | 15.512 | 18.221 | 23.423 | 31.092 | 36.519 | 46.944 | 155.569 | 182.720 | 234.877 |
| 3 | 5.788 | 6.823 | 8.819 | 8.306 | 9.789 | 12.647 | 18.782 | 22.131 | 28.586 |
| 4 | 4.157 | 4.913 | 6.372 | 5.368 | 6.341 | 8.221 | 9.416 | 11.118 | 14.405 |
| 5 | 3.499 | 4.142 | 5.387 | 4.291 | 5.077 | 6.598 | 6.655 | 7.870 | 10.220 |
| 6 | 3.141 | 3.723 | 4.850 | 3.733 | 4.422 | 5.758 | 5.383 | 6.376 | 8.292 |
| 7 | 2.913 | 3.456 | 4.508 | 3.390 | 4.020 | 5.241 | 4.658 | 5.520 | 7.191 |
| 8 | 2.754 | 3.270 | 4.271 | 3.156 | 3.746 | 4.889 | 4.189 | 4.968 | 6.479 |
| 9 | 2.637 | 3.132 | 4.094 | 2.986 | 3.546 | 4.633 | 3.860 | 4.581 | 5.980 |
| 10 | 2.546 | 3.026 | 3.958 | 2.856 | 3.393 | 4.437 | 3.617 | 4.294 | 5.610 |
| 12 | 2.414 | 2.871 | 3.759 | 2.670 | 3.175 | 4.156 | 3.279 | 3.896 | 5.096 |
| 14 | 2.322 | 2.762 | 3.620 | 2.542 | 3.024 | 3.962 | 3.054 | 3.631 | 4.753 |
| 16 | 2.254 | 2.682 | 3.517 | 2.449 | 2.913 | 3.819 | 2.89 | 3.441 | 4.507 |
| 18 | 2.201 | 2.620 | 3.436 | 2.376 | 2.828 | 3.709 | 2.771 | 3.297 | 4.321 |
| 20 | 2.158 | 2.570 | 3.372 | 2.319 | 2.760 | 3.621 | 2.675 | 3.184 | 4.175 |
| 22 | 2.123 | 2.528 | 3.318 | 2.272 | 2.705 | 3.549 | 2.598 | 3.092 | 4.056 |
| 24 | 2.094 | 2.494 | 3.274 | 2.232 | 2.658 | 3.489 | 2.534 | 3.017 | 3.958 |
| 26 | 2.069 | 2.464 | 3.235 | 2.199 | 2.619 | 3.437 | 2.480 | 2.953 | 3.875 |
| 28 | 2.048 | 2.439 | 3.202 | 2.170 | 2.585 | 3.393 | 2.434 | 2.898 | 3.804 |
| 30 | 2.029 | 2.417 | 3.173 | 2.145 | 2.555 | 3.355 | 2.394 | 2.851 | 3.742 |
| 35 | 1.991 | 2.371 | 3.114 | 2.094 | 2.495 | 3.276 | 2.314 | 2.756 | 3.618 |
| 40 | 1.961 | 2.336 | 3.069 | 2.055 | 2.448 | 3.216 | 2.253 | 2.684 | 3.524 |
| 45 | 1.938 | 2.308 | 3.032 | 2.024 | 2.412 | 3.168 | 2.205 | 2.627 | 3.450 |
| 50 | 1.918 | 2.285 | 3.003 | 1.999 | 2.382 | 3.129 | 2.166 | 2.580 | 3.390 |
| 60 | 1.888 | 2.250 | 2.956 | 1.960 | 2.335 | 3.068 | 2.106 | 2.509 | 3.297 |
| 70 | 1.866 | 2.224 | 2.922 | 1.931 | 2.300 | 3.023 | 2.062 | 2.457 | 3.228 |
| 80 | 1.849 | 2.203 | 2.895 | 1.908 | 2.274 | 2.988 | 2.028 | 2.416 | 3.175 |
| 90 | 1.835 | 2.186 | 2.873 | 1.890 | 2.252 | 2.959 | 2.001 | 2.384 | 3.133 |
| 100 | 1.823 | 2.172 | 2.855 | 1.875 | 2.234 | 2.936 | 1.978 | 2.357 | 3.098 |
| 150 | 1.786 | 2.128 | 2.796 | 1.826 | 2.176 | 2.859 | 1.906 | 2.271 | 2.985 |
| 200 | 1.764 | 2.102 | 2.763 | 1.798 | 2.143 | 2.816 | 1.866 | 2.223 | 2.921 |
| 250 | 1.750 | 2.085 | 2.741 | 1.780 | 2.121 | 2.788 | 1.839 | 2.191 | 2.880 |
| 300 | 1.740 | 2.073 | 2.725 | 1.767 | 2.106 | 2.767 | 1.820 | 2.169 | 2.850 |
| 350 | 1.732 | 2.064 | 2.713 | 1.757 | 2.094 | 2.752 | 1.806 | 2.152 | 2.828 |
| 400 | 1.726 | 2.057 | 2.703 | 1.749 | 2.084 | 2.739 | 1.794 | 2.138 | 2.810 |
| 500 | 1.717 | 2.046 | 2.689 | 1.737 | 2.070 | 2.721 | 1.777 | 2.117 | 2.783 |
| 1,000 | 1.695 | 2.019 | 2.654 | 1.709 | 2.036 | 2.676 | 1.736 | 2.068 | 2.718 |
| ∞ | 1.645 | 1.960 | 2.576 | 1.645 | 1.960 | 2.576 | 1.645 | 1.960 | 2.576 |

Adapted from Odeh, R. E., and D. B. Owen, 1980, *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, Marcel Dekker, Inc., New York, NY with permission.

Table T-11b: One-sided tolerance limit factors for a normal distribution
(see Section 9.12)

| n | $\gamma = 0.90$ | | | $\gamma = 0.95$ | | | $\gamma = 0.99$ | | |
|----------|-----------------|--------------|--------------|-----------------|--------------|--------------|-----------------|--------------|--------------|
| | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ | $\pi = 0.90$ | $\pi = 0.95$ | $\pi = 0.99$ |
| 2 | 10.253 | 13.090 | 18.500 | 20.581 | 26.260 | 37.094 | 103.029 | 131.426 | 185.617 |
| 3 | 4.258 | 5.311 | 7.340 | 6.155 | 7.656 | 10.553 | 13.995 | 17.370 | 23.896 |
| 4 | 3.188 | 3.957 | 5.438 | 4.162 | 5.144 | 7.042 | 7.380 | 9.083 | 12.387 |
| 5 | 2.742 | 3.400 | 4.666 | 3.407 | 4.203 | 5.741 | 5.362 | 6.578 | 8.939 |
| 6 | 2.494 | 3.092 | 4.243 | 3.006 | 3.708 | 5.062 | 4.411 | 5.406 | 7.335 |
| 7 | 2.333 | 2.894 | 3.972 | 2.755 | 3.399 | 4.642 | 3.859 | 4.728 | 6.412 |
| 8 | 2.219 | 2.754 | 3.783 | 2.582 | 3.187 | 4.354 | 3.497 | 4.285 | 5.812 |
| 9 | 2.133 | 2.650 | 3.641 | 2.454 | 3.031 | 4.143 | 3.240 | 3.972 | 5.389 |
| 10 | 2.066 | 2.568 | 3.532 | 2.355 | 2.911 | 3.981 | 3.048 | 3.738 | 5.074 |
| 12 | 1.966 | 2.448 | 3.371 | 2.210 | 2.736 | 3.747 | 2.777 | 3.410 | 4.633 |
| 14 | 1.895 | 2.363 | 3.257 | 2.109 | 2.614 | 3.585 | 2.596 | 3.189 | 4.337 |
| 16 | 1.842 | 2.299 | 3.172 | 2.033 | 2.524 | 3.464 | 2.459 | 3.028 | 4.123 |
| 18 | 1.800 | 2.249 | 3.105 | 1.974 | 2.453 | 3.370 | 2.357 | 2.905 | 3.960 |
| 20 | 1.765 | 2.208 | 3.052 | 1.926 | 2.396 | 3.295 | 2.276 | 2.808 | 3.832 |
| 22 | 1.737 | 2.174 | 3.007 | 1.886 | 2.349 | 3.233 | 2.209 | 2.729 | 3.727 |
| 24 | 1.712 | 2.145 | 2.969 | 1.853 | 2.309 | 3.181 | 2.154 | 2.662 | 3.640 |
| 26 | 1.691 | 2.120 | 2.937 | 1.824 | 2.275 | 3.136 | 2.106 | 2.606 | 3.566 |
| 28 | 1.673 | 2.099 | 2.909 | 1.799 | 2.246 | 3.098 | 2.065 | 2.558 | 3.502 |
| 30 | 1.657 | 2.080 | 2.884 | 1.777 | 2.220 | 3.064 | 2.030 | 2.515 | 3.447 |
| 35 | 1.624 | 2.041 | 2.833 | 1.732 | 2.167 | 2.995 | 1.957 | 2.430 | 3.334 |
| 40 | 1.598 | 2.010 | 2.793 | 1.697 | 2.125 | 2.941 | 1.902 | 2.364 | 3.249 |
| 45 | 1.577 | 1.986 | 2.761 | 1.669 | 2.092 | 2.898 | 1.857 | 2.312 | 3.180 |
| 50 | 1.559 | 1.965 | 2.735 | 1.646 | 2.065 | 2.862 | 1.821 | 2.269 | 3.125 |
| 60 | 1.532 | 1.933 | 2.694 | 1.609 | 2.022 | 2.807 | 1.764 | 2.202 | 3.038 |
| 70 | 1.511 | 1.909 | 2.662 | 1.581 | 1.990 | 2.765 | 1.722 | 2.153 | 2.974 |
| 80 | 1.495 | 1.890 | 2.638 | 1.559 | 1.964 | 2.733 | 1.688 | 2.114 | 2.924 |
| 90 | 1.481 | 1.874 | 2.618 | 1.542 | 1.944 | 2.706 | 1.661 | 2.082 | 2.883 |
| 100 | 1.470 | 1.861 | 2.601 | 1.527 | 1.927 | 2.684 | 1.639 | 2.056 | 2.850 |
| 150 | 1.433 | 1.818 | 2.546 | 1.478 | 1.870 | 2.611 | 1.566 | 1.971 | 2.740 |
| 200 | 1.411 | 1.793 | 2.514 | 1.450 | 1.837 | 2.570 | 1.524 | 1.923 | 2.679 |
| 250 | 1.397 | 1.777 | 2.493 | 1.431 | 1.815 | 2.542 | 1.496 | 1.891 | 2.638 |
| 300 | 1.386 | 1.765 | 2.477 | 1.417 | 1.800 | 2.522 | 1.475 | 1.868 | 2.608 |
| 350 | 1.378 | 1.755 | 2.466 | 1.406 | 1.787 | 2.506 | 1.461 | 1.850 | 2.585 |
| 400 | 1.372 | 1.748 | 2.456 | 1.398 | 1.778 | 2.494 | 1.448 | 1.836 | 2.567 |
| 500 | 1.362 | 1.736 | 2.442 | 1.385 | 1.763 | 2.475 | 1.430 | 1.814 | 2.540 |
| 1,000 | 1.338 | 1.709 | 2.407 | 1.354 | 1.727 | 2.430 | 1.385 | 1.762 | 2.475 |
| ∞ | 1.282 | 1.645 | 2.326 | 1.282 | 1.645 | 2.326 | 1.282 | 1.645 | 2.326 |

Adapted from Odeh, R. E., and D. B. Owen, 1980, *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, Marcel Dekker, Inc., New York, NY with permission.

Table T-12: Two thousand random digits

| | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 | 46-50 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 56580 | 93353 | 14830 | 88611 | 77491 | 94159 | 55682 | 61651 | 76278 | 90039 |
| 2 | 82694 | 44512 | 56868 | 96549 | 97676 | 81145 | 51299 | 78324 | 87458 | 19158 |
| 3 | 24300 | 89591 | 51838 | 78203 | 68168 | 64742 | 93755 | 56080 | 61768 | 16029 |
| 4 | 69791 | 64849 | 47370 | 41245 | 59336 | 91731 | 78722 | 07645 | 56793 | 21081 |
| 5 | 93678 | 10246 | 68835 | 27682 | 60318 | 07379 | 19157 | 12037 | 67776 | 07248 |
| 6 | 07473 | 88205 | 27403 | 13619 | 74578 | 16637 | 14566 | 40858 | 58759 | 24621 |
| 7 | 93255 | 94101 | 44641 | 53302 | 91743 | 92258 | 18179 | 07676 | 39974 | 51887 |
| 8 | 28910 | 26516 | 25706 | 48056 | 50645 | 72581 | 84289 | 31997 | 18444 | 10330 |
| 9 | 70419 | 62779 | 69059 | 74384 | 06975 | 92192 | 23660 | 33342 | 31465 | 83723 |
| 10 | 95145 | 92735 | 91272 | 57287 | 53395 | 79920 | 93979 | 99456 | 85530 | 79258 |
| 11 | 27054 | 06908 | 12502 | 66882 | 86941 | 37710 | 17092 | 53887 | 80494 | 09245 |
| 12 | 60999 | 30405 | 16956 | 04938 | 51713 | 78369 | 91595 | 37502 | 22165 | 91337 |
| 13 | 50513 | 14966 | 28936 | 51542 | 27276 | 40725 | 81142 | 10178 | 36861 | 32862 |
| 14 | 14918 | 14251 | 63794 | 60699 | 44952 | 74709 | 24733 | 50845 | 64032 | 26115 |
| 15 | 83143 | 72762 | 92850 | 34146 | 69102 | 21201 | 18647 | 69705 | 05843 | 24621 |
| 16 | 44762 | 01384 | 61844 | 28710 | 93492 | 04594 | 99063 | 72617 | 63939 | 59104 |
| 17 | 90525 | 13775 | 15676 | 30909 | 49632 | 93762 | 20406 | 59516 | 06116 | 45980 |
| 18 | 65288 | 57500 | 42177 | 82255 | 52023 | 99471 | 15693 | 20134 | 89639 | 96221 |
| 19 | 62309 | 36394 | 77163 | 92427 | 65271 | 89899 | 93288 | 15417 | 89998 | 13986 |
| 20 | 90434 | 39040 | 29549 | 48332 | 12172 | 32711 | 98444 | 67332 | 49853 | 76770 |
| 21 | 71192 | 17298 | 21629 | 28567 | 45628 | 99871 | 76063 | 03132 | 69163 | 92841 |
| 22 | 39191 | 99240 | 40029 | 32771 | 39050 | 95144 | 07049 | 36518 | 08289 | 92136 |
| 23 | 94055 | 84500 | 26272 | 27985 | 91858 | 60511 | 91802 | 13735 | 95525 | 04157 |
| 24 | 10718 | 72291 | 26193 | 79285 | 29209 | 87962 | 13485 | 53738 | 08642 | 22828 |
| 25 | 83134 | 62665 | 17823 | 13358 | 55677 | 84591 | 81232 | 50910 | 83995 | 93294 |
| 26 | 93392 | 93965 | 88188 | 82628 | 68956 | 39247 | 89597 | 40521 | 17850 | 20716 |
| 27 | 33193 | 11074 | 13299 | 70680 | 95082 | 67903 | 50078 | 55113 | 00057 | 29045 |
| 28 | 65846 | 08032 | 65737 | 45903 | 25773 | 17815 | 58805 | 48617 | 93974 | 06308 |
| 29 | 41773 | 34637 | 59350 | 31782 | 57933 | 15605 | 13648 | 93406 | 89275 | 97494 |
| 30 | 43908 | 04150 | 37788 | 53587 | 05483 | 53569 | 96579 | 52470 | 42818 | 62589 |
| 31 | 72367 | 74503 | 61417 | 67453 | 83903 | 43326 | 52742 | 77183 | 66826 | 28232 |
| 32 | 93703 | 43522 | 26709 | 43846 | 47587 | 45688 | 67433 | 15227 | 37428 | 37437 |
| 33 | 23622 | 70833 | 97297 | 07793 | 97656 | 24860 | 74883 | 77142 | 31451 | 12219 |
| 34 | 06461 | 72197 | 26511 | 74964 | 54490 | 67611 | 54022 | 16892 | 80799 | 65864 |
| 35 | 50175 | 29889 | 18630 | 20552 | 60114 | 20652 | 22881 | 79254 | 88917 | 85399 |
| 36 | 10942 | 67178 | 93214 | 09890 | 51054 | 29666 | 10428 | 96478 | 88082 | 41255 |
| 37 | 91159 | 14499 | 26915 | 90942 | 69215 | 23544 | 57355 | 11040 | 83574 | 10788 |
| 38 | 88368 | 26257 | 43969 | 58835 | 92663 | 54980 | 50501 | 32510 | 94197 | 51546 |
| 39 | 09587 | 55021 | 86738 | 51441 | 25297 | 44086 | 80012 | 07801 | 11944 | 44102 |
| 40 | 46036 | 67486 | 53662 | 41233 | 76645 | 30871 | 17030 | 86671 | 49669 | 01711 |

Table T-13: Quantiles, $d_q(n)$, of the Kolmogorov test statistic
(see Section 11.8)

| Two-sided test | | | | | | Two-sided test | | | | | |
|----------------------------|----------|--------|--------|--------|--------|---|-------------|-------------|-------------|-------------|--------|
| | $p=0.80$ | 0.90 | 0.95 | 0.98 | 0.99 | | $p=0.80$ | 0.90 | 0.95 | 0.98 | 0.99 |
| $n=1$ | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 | $n=21$ | 0.226 | 0.259 | 0.281 | 0.321 | 0.344 |
| 2 | 0.684 | 0.776 | 0.842 | 0.900 | 0.929 | 22 | 0.221 | 0.253 | 0.281 | 0.314 | 0.337 |
| 3 | 0.565 | 0.636 | 0.708 | 0.735 | 0.829 | 23 | 0.216 | 0.247 | 0.275 | 0.307 | 0.330 |
| 4 | 0.493 | 0.565 | 0.624 | 0.639 | 0.734 | 24 | 0.212 | 0.242 | 0.269 | 0.301 | 0.323 |
| 5 | 0.447 | 0.509 | 0.563 | 0.627 | 0.669 | 25 | 0.208 | 0.238 | 0.264 | 0.295 | 0.317 |
| 6 | 0.410 | 0.468 | 0.519 | 0.577 | 0.617 | 26 | 0.204 | 0.233 | 0.259 | 0.290 | 0.311 |
| 7 | 0.381 | 0.436 | 0.483 | 0.538 | 0.576 | 27 | 0.200 | 0.229 | 0.254 | 0.284 | 0.305 |
| 8 | 0.358 | 0.410 | 0.454 | 0.507 | 0.542 | 28 | 0.197 | 0.225 | 0.250 | 0.279 | 0.300 |
| 9 | 0.339 | 0.387 | 0.430 | 0.480 | 0.513 | 29 | 0.193 | 0.221 | 0.246 | 0.275 | 0.295 |
| 10 | 0.323 | 0.369 | 0.409 | 0.457 | 0.489 | 30 | 0.190 | 0.218 | 0.242 | 0.270 | 0.290 |
| 11 | 0.308 | 0.352 | 0.391 | 0.437 | 0.468 | 31 | 0.187 | 0.214 | 0.238 | 0.266 | 0.285 |
| 12 | 0.296 | 0.338 | 0.375 | 0.419 | 0.449 | 32 | 0.184 | 0.211 | 0.234 | 0.262 | 0.281 |
| 13 | 0.285 | 0.325 | 0.361 | 0.404 | 0.432 | 33 | 0.182 | 0.208 | 0.231 | 0.258 | 0.277 |
| 14 | 0.275 | 0.314 | 0.349 | 0.390 | 0.418 | 34 | 0.179 | 0.205 | 0.227 | 0.254 | 0.273 |
| 15 | 0.266 | 0.304 | 0.338 | 0.377 | 0.404 | 35 | 0.177 | 0.202 | 0.224 | 0.251 | 0.269 |
| 16 | 0.258 | 0.295 | 0.327 | 0.366 | 0.392 | 36 | 0.174 | 0.199 | 0.221 | 0.247 | 0.265 |
| 17 | 0.250 | 0.286 | 0.318 | 0.355 | 0.381 | 37 | 0.172 | 0.196 | 0.218 | 0.244 | 0.262 |
| 18 | 0.244 | 0.279 | 0.309 | 0.346 | 0.371 | 38 | 0.170 | 0.194 | 0.215 | 0.241 | 0.258 |
| 19 | 0.237 | 0.271 | 0.301 | 0.337 | 0.361 | 39 | 0.168 | 0.191 | 0.213 | 0.238 | 0.255 |
| 20 | 0.232 | 0.265 | 0.294 | 0.329 | 0.352 | 40 | 0.165 | 0.189 | 0.210 | 0.235 | 0.252 |
| | | | | | | $p=.80$ 0.90 0.95 0.98 0.99 | | | | | |
| Approximation for $n > 40$ | | | | | | <u>1.07</u> | <u>1.22</u> | <u>1.36</u> | <u>1.52</u> | <u>1.63</u> | |
| | | | | | | \sqrt{n} | \sqrt{n} | \sqrt{n} | \sqrt{n} | \sqrt{n} | |

Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from L. H. Miller, 1956, *Journal of the American Statistical Association*, 51: 111-121 (Appendix). Reprinted with the permission of John Wiley & Sons, Inc., and the American Statistical Association.

The entries in this table are selected quantiles $d_q(n)$ of the Kolmogorov test statistic D for the two-sided test. Reject H_0 at the level α if D exceeds the $1 - \alpha$ quantile given in this table. These quantiles are exact for $n \leq 40$. A better approximation for $n > 40$ results if $\sqrt{n + \sqrt{n}} / 10$ is used instead of \sqrt{n} in the denominator.

Table T-14: Quantiles, $D'_q(n)$, of the distribution of the D' statistic
(see Section 11.10)

| n | q | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
| 50 | 93.82 | 94.58 | 95.6 | 96.41 | 97.24 | 100.9 | 101.1 | 101.3 | 101.4 | 101.5 |
| 52 | 99.66 | 100.4 | 101.5 | 102.3 | 103.2 | 107.0 | 107.2 | 107.4 | 107.6 | 107.7 |
| 54 | 105.6 | 106.4 | 107.5 | 108.4 | 109.3 | 113.2 | 113.5 | 113.7 | 113.9 | 113.9 |
| 56 | 111.7 | 112.5 | 113.6 | 114.5 | 115.5 | 119.5 | 119.8 | 120 | 120.2 | 120.3 |
| 58 | 117.9 | 118.7 | 119.9 | 120.8 | 121.8 | 126.0 | 126.3 | 126.5 | 126.7 | 126.8 |
| 60 | 124.2 | 125.1 | 126.3 | 127.2 | 128.2 | 132.6 | 132.4 | 133.1 | 133.3 | 133.4 |
| 62 | 130.6 | 131.5 | 132.7 | 133.7 | 134.7 | 139.2 | 139.6 | 139.8 | 140.0 | 140.2 |
| 64 | 137.1 | 138.0 | 139.3 | 140.3 | 141.3 | 146.0 | 146.4 | 146.6 | 146.9 | 147.0 |
| 66 | 143.7 | 144.7 | 146.0 | 147.0 | 148.0 | 152.9 | 153.3 | 153.5 | 153.8 | 153.9 |
| 68 | 150.5 | 151.4 | 152.8 | 153.8 | 154.9 | 159.9 | 160.3 | 160.6 | 160.8 | 161.0 |
| 70 | 157.3 | 158.3 | 159.6 | 160.7 | 161.8 | 167.0 | 167.4 | 167.7 | 168.0 | 168.1 |
| 72 | 164.3 | 165.3 | 166.6 | 167.7 | 168.9 | 174.2 | 174.6 | 174.9 | 175.2 | 175.4 |
| 74 | 171.3 | 172.3 | 173.7 | 174.8 | 176.0 | 181.5 | 181.9 | 182.2 | 182.5 | 182.7 |
| 76 | 178.4 | 179.5 | 180.9 | 182.0 | 183.3 | 188.9 | 189.3 | 189.7 | 190.0 | 190.2 |
| 78 | 185.7 | 186.7 | 188.2 | 189.4 | 190.6 | 196.4 | 196.8 | 197.2 | 197.5 | 197.7 |
| 80 | 193.0 | 194.1 | 195.6 | 196.8 | 198.0 | 204.0 | 204.4 | 204.8 | 205.1 | 205.4 |
| 82 | 200.4 | 201.5 | 203.1 | 204.3 | 205.6 | 211.6 | 212.1 | 212.5 | 212.9 | 213.1 |
| 84 | 207.9 | 209.1 | 210.6 | 211.9 | 213.2 | 219.4 | 219.9 | 220.3 | 220.7 | 220.9 |
| 86 | 215.5 | 216.7 | 218.3 | 219.6 | 220.9 | 227.3 | 227.8 | 228.2 | 228.6 | 228.8 |
| 88 | 223.3 | 224.4 | 226.1 | 227.3 | 228.7 | 235.2 | 235.8 | 236.2 | 236.6 | 236.8 |
| 90 | 231.1 | 232.2 | 233.9 | 235.2 | 236.6 | 243.3 | 243.8 | 244.3 | 244.7 | 245.0 |
| 92 | 238.9 | 240.2 | 241.8 | 243.2 | 244.6 | 251.4 | 252.0 | 252.4 | 252.9 | 253.2 |
| 94 | 246.9 | 248.1 | 249.9 | 251.2 | 252.7 | 259.6 | 260.2 | 260.7 | 261.2 | 261.4 |
| 96 | 255.0 | 256.2 | 258.0 | 259.3 | 260.8 | 268.0 | 268.6 | 269.1 | 269.5 | 269.8 |
| 98 | 263.1 | 264.4 | 266.2 | 267.6 | 269.1 | 276.4 | 277.0 | 277.5 | 278.0 | 278.3 |
| 100 | 271.4 | 272.7 | 274.4 | 275.9 | 277.4 | 284.9 | 285.5 | 286.0 | 286.5 | 286.8 |
| 120 | 358.2 | 359.7 | 361.8 | 363.5 | 365.3 | 374.2 | 375.1 | 375.7 | 376.4 | 376.8 |
| 140 | 452.9 | 454.6 | 456.9 | 458.8 | 460.9 | 471.4 | 472.4 | 473.2 | 474.0 | 474.5 |
| 160 | 554.7 | 556.6 | 559.2 | 561.4 | 563.7 | 575.7 | 576.9 | 577.8 | 578.8 | 579.4 |
| 180 | 663.2 | 665.3 | 668.2 | 670.6 | 673.1 | 686.7 | 688.1 | 689.2 | 690.3 | 691.0 |
| 200 | 778.1 | 780.4 | 783.6 | 786.1 | 788.9 | 804.1 | 805.6 | 806.9 | 808.2 | 809.0 |
| 220 | 899.0 | 901.4 | 904.9 | 907.7 | 910.7 | 927.4 | 929.1 | 930.5 | 932.0 | 933.0 |
| 240 | 1026 | 1028 | 1032 | 1035 | 1038 | 1056 | 1058 | 1060 | 1062 | 1063 |
| 260 | 1158 | 1160 | 1164 | 1168 | 1171 | 1191 | 1193 | 1195 | 1197 | 1198 |
| 280 | 1295 | 1298 | 1302 | 1306 | 1309 | 1331 | 1333 | 1335 | 1337 | 1338 |
| 300 | 1437 | 1440 | 1445 | 1449 | 1453 | 1475 | 1478 | 1480 | 1482 | 1484 |
| 320 | 1585 | 1588 | 1593 | 1597 | 1601 | 1625 | 628 | 1630 | 1632 | 1634 |
| 340 | 1737 | 1740 | 1745 | 1749 | 1754 | 1780 | 1783 | 1785 | 1787 | 1789 |
| 360 | 1893 | 1897 | 1902 | 1907 | 1911 | 1939 | 1942 | 1944 | 1947 | 1949 |
| 380 | 2054 | 2058 | 2064 | 2068 | 2070 | 2102 | 2105 | 2108 | 2111 | 2113 |
| 400 | 2220 | 2224 | 2230 | 2234 | 2240 | 2270 | 2274 | 2276 | 2279 | 2281 |
| 420 | 2389 | 2394 | 2400 | 2405 | 2410 | 2442 | 2446 | 2449 | 2452 | 2454 |

This material is reproduced from ANSI N15.15-1974(R1981) with permission of the American National Standards Institute (ANSI) and the Institute of Nuclear Materials Management (INMM).

Table T-14: Quantiles, $D'_q(n)$, of the distribution of the D' statistic (Cont'd)

| n | q | | | | | | | | | |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
| 440 | 2563 | 2568 | 2574 | 2579 | 2585 | 2618 | 2622 | 2625 | 2629 | 2631 |
| 460 | 2741 | 2746 | 2752 | 2758 | 2763 | 2799 | 2803 | 2806 | 2810 | 2812 |
| 480 | 2923 | 2928 | 2934 | 2940 | 2946 | 2983 | 2987 | 2991 | 2994 | 2997 |
| 500 | 3108 | 3113 | 3120 | 3126 | 3133 | 3171 | 3175 | 3179 | 3183 | 3186 |
| 520 | 3298 | 3303 | 3310 | 3316 | 3323 | 3363 | 3367 | 3371 | 3375 | 3378 |
| 540 | 3491 | 3496 | 3504 | 3510 | 3517 | 3558 | 3563 | 3567 | 3572 | 3574 |
| 560 | 3688 | 3693 | 3701 | 3707 | 3715 | 3757 | 3762 | 3767 | 3771 | 3774 |
| 580 | 3888 | 3894 | 3902 | 3908 | 3916 | 3960 | 3965 | 3970 | 3975 | 3978 |
| 600 | 4092 | 4098 | 4106 | 4113 | 4121 | 4166 | 4172 | 4176 | 4181 | 4185 |
| 620 | 4299 | 4305 | 4314 | 4321 | 4329 | 4376 | 4382 | 4387 | 4392 | 4395 |
| 640 | 4510 | 4516 | 4525 | 4532 | 4540 | 4589 | 4595 | 4600 | 4605 | 4609 |
| 660 | 4724 | 4730 | 4739 | 4747 | 4755 | 4806 | 4812 | 4817 | 4823 | 4826 |
| 680 | 4941 | 4948 | 4957 | 4965 | 4974 | 5026 | 5032 | 5037 | 5043 | 5047 |
| 700 | 5162 | 5169 | 5178 | 5186 | 5195 | 5249 | 5255 | 5260 | 5266 | 5270 |
| 720 | 5386 | 5393 | 5403 | 5411 | 5420 | 5475 | 5482 | 5487 | 5493 | 5497 |
| 740 | 5613 | 5620 | 5630 | 5638 | 5648 | 5704 | 5711 | 5717 | 5723 | 5727 |
| 760 | 5843 | 5850 | 5861 | 5869 | 5879 | 5937 | 5944 | 5950 | 5956 | 5961 |
| 780 | 6076 | 6084 | 6094 | 6103 | 6113 | 6172 | 6180 | 6186 | 6192 | 6197 |
| 800 | 6312 | 6320 | 6331 | 6340 | 6350 | 6411 | 6418 | 6425 | 6432 | 6436 |
| 850 | 6916 | 6924 | 6935 | 6945 | 6955 | 7020 | 7028 | 7035 | 7042 | 7047 |
| 900 | 7538 | 7546 | 7558 | 7568 | 7579 | 7648 | 7656 | 7664 | 7672 | 7677 |
| 950 | 8177 | 8186 | 8198 | 8209 | 8221 | 8293 | 8302 | 8310 | 8318 | 8324 |
| 1000 | 8834 | 8843 | 8856 | 8867 | 8879 | 8956 | 8965 | 8973 | 8982 | 8988 |
| 1050 | 9507 | 9516 | 9530 | 9542 | 9555 | 9635 | 9645 | 9653 | 9663 | 9669 |
| 1100 | 10200 | 10210 | 10220 | 10230 | 10250 | 10330 | 10340 | 10350 | 10360 | 10370 |
| 1150 | 10900 | 10910 | 10930 | 10940 | 10950 | 11040 | 11050 | 11060 | 11070 | 11080 |
| 1200 | 11620 | 11630 | 11650 | 11660 | 11680 | 11770 | 11780 | 11790 | 11800 | 11810 |
| 1250 | 12360 | 12370 | 12390 | 17400 | 12420 | 12510 | 12520 | 12530 | 12540 | 12550 |
| 1300 | 13110 | 13120 | 13140 | 13150 | 13170 | 13270 | 13280 | 13290 | 13300 | 13310 |
| 1350 | 13880 | 13890 | 13910 | 13920 | 13940 | 14040 | 14050 | 14060 | 14080 | 14090 |
| 1400 | 14660 | 14670 | 14690 | 14700 | 14720 | 14830 | 14840 | 14850 | 14860 | 14870 |
| 1450 | 15450 | 15460 | 15480 | 15500 | 15520 | 15630 | 15640 | 15650 | 15670 | 15680 |
| 1500 | 16260 | 16270 | 16290 | 16310 | 16330 | 16440 | 16460 | 16470 | 16480 | 16490 |
| 1550 | 17080 | 17100 | 17120 | 17130 | 17150 | 17270 | 17280 | 17300 | 17310 | 17320 |
| 1600 | 17920 | 17930 | 17950 | 17970 | 17990 | 18110 | 18130 | 18140 | 18160 | 18170 |
| 1650 | 18770 | 18780 | 18800 | 18820 | 18840 | 18970 | 18980 | 19000 | 19010 | 19020 |
| 1700 | 19630 | 19640 | 19670 | 19680 | 19700 | 19830 | 19850 | 19860 | 19880 | 19890 |
| 1750 | 20500 | 20520 | 20540 | 20560 | 20580 | 20710 | 20730 | 20750 | 20760 | 20770 |
| 1800 | 21390 | 21410 | 21430 | 21450 | 21470 | 21610 | 21630 | 21640 | 21660 | 21670 |
| 1850 | 22290 | 22310 | 22330 | 22350 | 22370 | 22510 | 22530 | 22550 | 22560 | 22580 |
| 1900 | 23200 | 23220 | 23240 | 23260 | 23290 | 23430 | 23450 | 23470 | 23480 | 23500 |
| 1950 | 24130 | 24140 | 24170 | 24190 | 24210 | 24360 | 24380 | 24400 | 24420 | 24430 |
| 2000 | 25060 | 25080 | 25110 | 25130 | 25150 | 25300 | 25320 | 25340 | 25360 | 25370 |

Table T-15: Cumulative probabilities $Pr\{U \leq u\}$ of the number of runs u in a sample of size $n_1 + n_2$, where n_1 is the number of plus signs and n_2 is the number of minus signs (see Section 25.3)

| (n_1, n_2) | u | | | | | | | | | |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| (2, 3) | 0.200 | 0.500 | 0.900 | 1.000 | | | | | | |
| (2, 4) | 0.133 | 0.400 | 0.800 | 1.000 | | | | | | |
| (2, 5) | 0.095 | 0.333 | 0.714 | 1.000 | | | | | | |
| (2, 6) | 0.071 | 0.286 | 0.643 | 1.000 | | | | | | |
| (2, 7) | 0.056 | 0.250 | 0.583 | 1.000 | | | | | | |
| (2, 8) | 0.044 | 0.222 | 0.533 | 1.000 | | | | | | |
| (2, 9) | 0.036 | 0.200 | 0.491 | 1.000 | | | | | | |
| (2, 10) | 0.030 | 0.182 | 0.455 | 1.000 | | | | | | |
| (3, 3) | 0.100 | 0.300 | 0.700 | 0.900 | 1.000 | | | | | |
| (3, 4) | 0.057 | 0.200 | 0.543 | 0.800 | 0.971 | 1.000 | | | | |
| (3, 5) | 0.036 | 0.143 | 0.429 | 0.714 | 0.929 | 1.000 | | | | |
| (3, 6) | 0.024 | 0.107 | 0.345 | 0.643 | 0.881 | 1.000 | | | | |
| (3, 7) | 0.017 | 0.083 | 0.283 | 0.583 | 0.833 | 1.000 | | | | |
| (3, 8) | 0.012 | 0.067 | 0.236 | 0.533 | 0.788 | 1.000 | | | | |
| (3, 9) | 0.009 | 0.055 | 0.200 | 0.491 | 0.745 | 1.000 | | | | |
| (3, 10) | 0.007 | 0.045 | 0.171 | 0.455 | 0.706 | 1.000 | | | | |
| (4, 4) | 0.029 | 0.114 | 0.371 | 0.629 | 0.886 | 0.971 | 1.000 | | | |
| (4, 5) | 0.016 | 0.071 | 0.262 | 0.500 | 0.786 | 0.929 | 0.992 | 1.000 | | |
| (4, 6) | 0.010 | 0.048 | 0.190 | 0.405 | 0.690 | 0.881 | 0.976 | 1.000 | | |
| (4, 7) | 0.006 | 0.033 | 0.142 | 0.333 | 0.606 | 0.833 | 0.954 | 1.000 | | |
| (4, 8) | 0.004 | 0.024 | 0.109 | 0.279 | 0.533 | 0.788 | 0.929 | 1.000 | | |
| (4, 9) | 0.003 | 0.018 | 0.085 | 0.236 | 0.471 | 0.745 | 0.902 | 1.000 | | |
| (4, 10) | 0.002 | 0.014 | 0.068 | 0.203 | 0.419 | 0.706 | 0.874 | 1.000 | | |
| (5, 5) | 0.008 | 0.040 | 0.167 | 0.357 | 0.643 | 0.833 | 0.960 | 0.992 | 1.000 | |
| (5, 6) | 0.004 | 0.024 | 0.110 | 0.262 | 0.522 | 0.738 | 0.911 | 0.976 | 0.998 | |
| (5, 7) | 0.003 | 0.015 | 0.076 | 0.197 | 0.424 | 0.652 | 0.854 | 0.955 | 0.992 | |
| (5, 8) | 0.002 | 0.010 | 0.054 | 0.152 | 0.347 | 0.576 | 0.793 | 0.929 | 0.984 | |
| (5, 9) | 0.001 | 0.007 | 0.039 | 0.119 | 0.287 | 0.510 | 0.734 | 0.902 | 0.972 | |
| (5, 10) | 0.001 | 0.005 | 0.029 | 0.095 | 0.239 | 0.455 | 0.678 | 0.874 | 0.958 | |
| (6, 6) | 0.002 | 0.013 | 0.067 | 0.175 | 0.392 | 0.608 | 0.825 | 0.933 | 0.987 | |
| (6, 7) | 0.001 | 0.008 | 0.043 | 0.121 | 0.296 | 0.500 | 0.733 | 0.879 | 0.966 | |
| (6, 8) | 0.001 | 0.005 | 0.028 | 0.086 | 0.226 | 0.413 | 0.646 | 0.821 | 0.937 | |
| (6, 9) | 0.000 | 0.003 | 0.019 | 0.063 | 0.175 | 0.343 | 0.566 | 0.762 | 0.902 | |
| (6, 10) | 0.000 | 0.002 | 0.013 | 0.047 | 0.137 | 0.288 | 0.497 | 0.706 | 0.864 | |
| (7, 7) | 0.001 | 0.004 | 0.025 | 0.078 | 0.209 | 0.383 | 0.617 | 0.791 | 0.922 | |
| (7, 8) | 0.000 | 0.002 | 0.015 | 0.051 | 0.149 | 0.296 | 0.514 | 0.704 | 0.867 | |
| (7, 9) | 0.000 | 0.001 | 0.010 | 0.035 | 0.108 | 0.231 | 0.427 | 0.622 | 0.806 | |
| (7, 10) | 0.000 | 0.001 | 0.006 | 0.024 | 0.080 | 0.182 | 0.355 | 0.549 | 0.743 | |
| (8, 8) | 0.000 | 0.001 | 0.009 | 0.032 | 0.100 | 0.214 | 0.405 | 0.595 | 0.786 | |
| (8, 9) | 0.000 | 0.001 | 0.005 | 0.022 | 0.069 | 0.157 | 0.319 | 0.500 | 0.702 | |
| (8, 10) | 0.000 | 0.000 | 0.003 | 0.013 | 0.048 | 0.117 | 0.251 | 0.419 | 0.621 | |
| (9, 9) | 0.000 | 0.000 | 0.003 | 0.012 | 0.044 | 0.109 | 0.238 | 0.399 | 0.601 | |
| (9, 10) | 0.000 | 0.000 | 0.002 | 0.008 | 0.029 | 0.077 | 0.179 | 0.319 | 0.510 | |
| (10, 10) | 0.000 | 0.000 | 0.001 | 0.004 | 0.019 | 0.051 | 0.128 | 0.242 | 0.414 | |

Table T-15: Cumulative probabilities $Pr\{U \leq u\}$ of the number of runs u in a sample of size $n_1 + n_2$ (Continued)

| (n_1, n_2) | u | | | | | | | | | |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| (5, 6) | 1.000 | | | | | | | | | |
| (5, 7) | 1.000 | | | | | | | | | |
| (5, 8) | 1.000 | | | | | | | | | |
| (5, 9) | 1.000 | | | | | | | | | |
| (5, 10) | 1.000 | | | | | | | | | |
| (6, 6) | 0.998 | 1.000 | | | | | | | | |
| (6, 7) | 0.992 | 0.999 | 1.000 | | | | | | | |
| (6, 8) | 0.984 | 0.998 | 1.000 | | | | | | | |
| (6, 9) | 0.972 | 0.994 | 1.000 | | | | | | | |
| (6, 10) | 0.958 | 0.990 | 1.000 | | | | | | | |
| (7, 7) | 0.975 | 0.996 | 0.999 | 1.000 | | | | | | |
| (7, 8) | 0.949 | 0.988 | 0.998 | 1.000 | 1.000 | | | | | |
| (7, 9) | 0.916 | 0.975 | 0.994 | 0.999 | 1.000 | | | | | |
| (7, 10) | 0.879 | 0.957 | 0.990 | 0.998 | 1.000 | | | | | |
| (8, 8) | 0.900 | 0.968 | 0.991 | 0.999 | 1.000 | 1.000 | | | | |
| (8, 9) | 0.843 | 0.939 | 0.980 | 0.996 | 0.999 | 1.000 | 1.000 | | | |
| (8, 10) | 0.782 | 0.903 | 0.964 | 0.990 | 0.998 | 1.000 | 1.000 | | | |
| (9, 9) | 0.762 | 0.891 | 0.956 | 0.988 | 0.997 | 1.000 | 1.000 | 1.000 | | |
| (9, 10) | 0.681 | 0.834 | 0.923 | 0.974 | 0.992 | 0.999 | 1.000 | 1.000 | 1.000 | |
| (10, 10) | 0.586 | 0.758 | 0.872 | 0.949 | 0.981 | 0.996 | 0.999 | 1.000 | 1.000 | 1.000 |

Reprinted with the permission of the Institute of Mathematical Statistics.

- Although probabilities are given for (n_1, n_2) with $n_1 \leq n_2$, the table is symmetric, and can be used for (n_1, n_2) with $(n_1 > n_2)$

Table T-16: Quantiles, W_q , of the Wilcoxon signed ranks test statistic
(see Section 25.5)

| | $w_{0.005}$ | $w_{0.01}$ | $w_{0.025}$ | $w_{0.05}$ | $w_{0.10}$ | $w_{0.20}$ | $w_{0.30}$ | $w_{0.40}$ | $w_{0.50}$ | $n(n+1)/2$ |
|---------------------------|-------------|------------|-------------|------------|------------|------------|------------|------------|------------|------------|
| $n = 4$ | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 4 | 5 | 10 |
| 5 | 0 | 0 | 0 | 1 | 3 | 4 | 5 | 6 | 7.5 | 15 |
| 6 | 0 | 0 | 1 | 3 | 4 | 6 | 8 | 9 | 10.5 | 21 |
| 7 | 0 | 1 | 3 | 4 | 6 | 9 | 11 | 12 | 14 | 28 |
| 8 | 1 | 2 | 4 | 6 | 9 | 12 | 14 | 16 | 18 | 36 |
| 9 | 2 | 4 | 6 | 9 | 11 | 15 | 18 | 20 | 22.5 | 45 |
| 10 | 4 | 6 | 9 | 11 | 15 | 19 | 22 | 25 | 27.5 | 55 |
| 11 | 6 | 8 | 11 | 14 | 18 | 23 | 27 | 30 | 33 | 66 |
| 12 | 8 | 10 | 14 | 18 | 22 | 28 | 32 | 36 | 39 | 78 |
| 13 | 10 | 13 | 18 | 22 | 27 | 33 | 38 | 42 | 45.5 | 91 |
| 14 | 13 | 16 | 22 | 26 | 32 | 39 | 44 | 48 | 52.5 | 105 |
| 15 | 16 | 20 | 26 | 31 | 37 | 45 | 51 | 55 | 60 | 120 |
| 16 | 20 | 24 | 30 | 36 | 43 | 51 | 58 | 63 | 68 | 136 |
| 17 | 24 | 28 | 35 | 42 | 49 | 58 | 65 | 71 | 76.5 | 153 |
| 18 | 28 | 33 | 41 | 48 | 56 | 66 | 73 | 80 | 85.15 | 171 |
| 19 | 33 | 38 | 47 | 54 | 63 | 74 | 82 | 89 | 95 | 190 |
| 20 | 38 | 44 | 53 | 61 | 70 | 83 | 91 | 98 | 105 | 210 |
| 21 | 44 | 50 | 59 | 68 | 78 | 91 | 100 | 108 | 115.5 | 231 |
| 22 | 49 | 56 | 67 | 76 | 87 | 100 | 110 | 119 | 126.5 | 253 |
| 23 | 55 | 63 | 74 | 84 | 95 | 110 | 120 | 130 | 138 | 276 |
| 24 | 62 | 70 | 82 | 92 | 105 | 120 | 131 | 141 | 150 | 300 |
| 25 | 69 | 77 | 90 | 101 | 114 | 131 | 143 | 153 | 162.5 | 325 |
| 26 | 76 | 85 | 99 | 111 | 125 | 142 | 155 | 165 | 175.5 | 351 |
| 27 | 84 | 94 | 108 | 120 | 135 | 154 | 167 | 178 | 189 | 378 |
| 28 | 92 | 102 | 117 | 131 | 146 | 166 | 180 | 192 | 203 | 406 |
| 29 | 101 | 111 | 127 | 141 | 158 | 178 | 193 | 206 | 217.5 | 435 |
| 30 | 110 | 121 | 138 | 152 | 170 | 191 | 207 | 220 | 232.5 | 465 |
| 31 | 119 | 131 | 148 | 164 | 182 | 205 | 221 | 235 | 248 | 496 |
| 32 | 129 | 141 | 160 | 176 | 195 | 219 | 236 | 250 | 264 | 528 |
| 33 | 139 | 152 | 171 | 188 | 208 | 233 | 251 | 266 | 280.5 | 561 |
| 34 | 149 | 163 | 183 | 201 | 222 | 248 | 266 | 282 | 297.5 | 595 |
| 35 | 160 | 175 | 196 | 214 | 236 | 263 | 283 | 299 | 315 | 630 |
| 36 | 172 | 187 | 209 | 228 | 251 | 279 | 299 | 317 | 333 | 666 |
| 37 | 184 | 199 | 222 | 242 | 266 | 295 | 316 | 335 | 351.5 | 703 |
| 38 | 196 | 212 | 236 | 257 | 282 | 312 | 334 | 353 | 370.5 | 741 |
| 39 | 208 | 225 | 250 | 272 | 298 | 329 | 352 | 372 | 390 | 780 |
| 40 | 221 | 239 | 265 | 287 | 314 | 347 | 371 | 391 | 410 | 820 |
| 41 | 235 | 253 | 280 | 303 | 331 | 365 | 390 | 411 | 430.5 | 861 |
| 42 | 248 | 267 | 295 | 320 | 349 | 384 | 409 | 431 | 451.5 | 903 |
| 43 | 263 | 282 | 311 | 337 | 366 | 403 | 429 | 452 | 473 | 946 |
| 44 | 277 | 297 | 328 | 354 | 385 | 422 | 450 | 473 | 495 | 990 |
| 45 | 292 | 313 | 344 | 372 | 403 | 442 | 471 | 495 | 517.5 | 1035 |
| 46 | 308 | 329 | 362 | 390 | 423 | 463 | 492 | 517 | 540.5 | 1081 |
| 47 | 324 | 346 | 379 | 408 | 442 | 484 | 514 | 540 | 564 | 1128 |
| 48 | 340 | 363 | 397 | 428 | 463 | 505 | 536 | 563 | 588 | 1176 |
| 49 | 357 | 381 | 416 | 447 | 483 | 527 | 559 | 587 | 612.5 | 1225 |
| 50 | 374 | 398 | 435 | 467 | 504 | 550 | 583 | 611 | 637.5 | 1275 |

Reprinted with the permission of the American Statistical Association

Table T-17: Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum (WRS) Statistic
(see Section 25.8)

| <i>n</i> | w_q | <i>m</i> | | | | | | | | | | | | | | | | | | |
|----------|-------|----------|----|----|----|----|----|----|----|----|----|----|----|-----|----|----|----|----|----|----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 2 | 0.001 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| | 0.005 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 |
| | 0.01 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| | 0.025 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 |
| | 0.05 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 |
| | 0.1 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 |
| 3 | 0.001 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 7 |
| | 0.005 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 10 | 10 |
| | 0.01 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 |
| | 0.025 | 6 | 6 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 14 | 14 | 15 |
| | 0.05 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 11 | 11 | 12 | 12 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 18 |
| | 0.1 | 7 | 8 | 8 | 9 | 10 | 11 | 12 | 12 | 13 | 14 | 15 | 16 | 17 | 17 | 18 | 19 | 20 | 21 | 22 |
| 4 | 0.001 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 12 | 13 | 13 | 14 | 14 | 14 | 14 |
| | 0.005 | 10 | 10 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 17 | 18 | 19 |
| | 0.01 | 10 | 10 | 10 | 11 | 12 | 12 | 13 | 14 | 14 | 15 | 16 | 16 | 17 | 18 | 18 | 19 | 20 | 20 | 21 |
| | 0.025 | 10 | 10 | 11 | 12 | 13 | 14 | 15 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 22 | 23 | 24 | 25 |
| | 0.05 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 25 | 26 | 27 | 28 | 29 |
| | 0.1 | 11 | 12 | 14 | 15 | 16 | 17 | 18 | 20 | 21 | 22 | 23 | 24 | 26 | 27 | 28 | 29 | 31 | 32 | 33 |
| 5 | 0.001 | 15 | 15 | 15 | 15 | 15 | 16 | 17 | 17 | 18 | 18 | 19 | 19 | 20 | 21 | 21 | 22 | 23 | 23 | |
| | 0.005 | 15 | 15 | 15 | 16 | 17 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| | 0.01 | 15 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| | 0.025 | 15 | 16 | 17 | 18 | 19 | 21 | 22 | 23 | 24 | 25 | 27 | 28 | 29 | 30 | 31 | 33 | 34 | 35 | 36 |
| | 0.05 | 16 | 17 | 18 | 20 | 21 | 22 | 24 | 25 | 27 | 28 | 29 | 31 | 32 | 34 | 35 | 36 | 38 | 39 | 41 |
| | 0.1 | 17 | 18 | 20 | 21 | 23 | 24 | 26 | 28 | 29 | 31 | 33 | 34 | 36 | 38 | 39 | 41 | 43 | 44 | 46 |
| 6 | 0.001 | 21 | 21 | 21 | 21 | 21 | 23 | 24 | 25 | 26 | 26 | 27 | 28 | -29 | 30 | 31 | 32 | 33 | 34 | |
| | 0.005 | 21 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 31 | 32 | 33 | 34 | 35 | 37 | 38 | 39 | 40 |
| | 0.01 | 21 | 21 | 23 | 24 | 25 | 26 | 28 | 29 | 30 | 31 | 33 | 34 | 35 | 37 | 38 | 40 | 41 | 42 | 44 |
| | 0.025 | 21 | 23 | 24 | 25 | 27 | 28 | 30 | 32 | 33 | 35 | 36 | 38 | 39 | 41 | 43 | 44 | 46 | 47 | 49 |
| | 0.05 | 22 | 24 | 25 | 27 | 29 | 30 | 32 | 34 | 36 | 38 | 39 | 41 | 43 | 45 | 47 | 48 | 50 | 52 | 54 |
| | 0.1 | 23 | 25 | 27 | 29 | 31 | 33 | 35 | 37 | 39 | 41 | 43 | 45 | 47 | 49 | 51 | 53 | 56 | 58 | 60 |

Table T-17: Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum(WRS) Statistic
(Continued)

| <i>n</i> | W_q | <i>m</i> | | | | | | | | | | | | | | | | | | |
|-----------|--------------|----------|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 7 | 0.001 | 28 | 28 | 28 | 28 | 29 | 30 | 31 | 32 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 42 | 43 | 44 | 45 |
| | 0.005 | 28 | 28 | 29 | 30 | 32 | 33 | 35 | 36 | 38 | 39 | 41 | 42 | 44 | 45 | 47 | 48 | 50 | 51 | 53 |
| | 0.01 | 28 | 29 | 30 | 32 | 33 | 35 | 36 | 38 | 40 | 41 | 43 | 45 | 46 | 48 | 50 | 52 | 53 | 55 | 57 |
| | 0.025 | 28 | 30 | 32 | 34 | 35 | 37 | 39 | 41 | 43 | 45 | 47 | 49 | 51 | 53 | 55 | 57 | 59 | 61 | 63 |
| | 0.05 | 29 | 31 | 33 | 35 | 37 | 40 | 42 | 44 | 46 | 48 | 50 | 53 | 55 | 57 | 59 | 62 | 64 | 66 | 68 |
| | 0.1 | 30 | 33 | 35 | 37 | 40 | 42 | 45 | 47 | 50 | 52 | 55 | 57 | 60 | 62 | 65 | 67 | 70 | 72 | 75 |
| 8 | 0.001 | 36 | 36 | 36 | 37 | 38 | 39 | 41 | 42 | 43 | 45 | 46 | 48 | 49 | 51 | 52 | 54 | 55 | 57 | 58 |
| | 0.005 | 36 | 36 | 38 | 39 | 41 | 43 | 44 | 46 | 48 | 50 | 52 | 54 | 55 | 57 | 59 | 61 | 63 | 65 | 67 |
| | 0.01 | 36 | 37 | 39 | 41 | 43 | 44 | 46 | 48 | 50 | 52 | 54 | 56 | 59 | 61 | 63 | 65 | 67 | 69 | 71 |
| | 0.025 | 37 | 39 | 41 | 43 | 45 | 47 | 50 | 52 | 54 | 56 | 59 | 61 | 63 | 66 | 68 | 71 | 73 | 75 | 78 |
| | 0.05 | 38 | 40 | 42 | 45 | 47 | 50 | 52 | 55 | 57 | 60 | 63 | 65 | 68 | 70 | 73 | 76 | 78 | 81 | 84 |
| | 0.1 | 39 | 42 | 44 | 47 | 50 | 53 | 56 | 59 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 88 | 91 |
| 9 | 0.001 | 45 | 45 | 45 | 47 | 48 | 49 | 51 | 53 | 54 | 56 | 58 | 60 | 61 | 63 | 65 | 67 | 69 | 71 | 72 |
| | 0.005 | 45 | 46 | 47 | 49 | 51 | 53 | 55 | 57 | 59 | 62 | 64 | 66 | 68 | 70 | 73 | 75 | 77 | 79 | 82 |
| | 0.01 | 45 | 47 | 49 | 51 | 53 | 55 | 57 | 60 | 62 | 64 | 67 | 69 | 72 | 74 | 77 | 79 | 82 | 84 | 86 |
| | 0.025 | 46 | 48 | 50 | 53 | 56 | 58 | 61 | 63 | 66 | 69 | 72 | 74 | 77 | 80 | 83 | 85 | 88 | 91 | 94 |
| | 0.05 | 47 | 50 | 52 | 55 | 58 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 88 | 91 | 94 | 97 | 100 |
| | 0.1 | 48 | 51 | 55 | 58 | 61 | 64 | 68 | 71 | 74 | 77 | 81 | 84 | 87 | 91 | 94 | 98 | 101 | 104 | 108 |
| 10 | 0.001 | 55 | 55 | 56 | 57 | 59 | 61 | 62 | 64 | 66 | 68 | 70 | 73 | 75 | 77 | 79 | 81 | 83 | 85 | 88 |
| | 0.005 | 55 | 56 | 58 | 60 | 62 | 65 | 67 | 69 | 72 | 74 | 77 | 80 | 82 | 85 | 87 | 90 | 93 | 95 | 98 |
| | 0.01 | 55 | 57 | 59 | 62 | 64 | 67 | 69 | 72 | 75 | 78 | 80 | 83 | 86 | 89 | 92 | 94 | 97 | 100 | 103 |
| | 0.025 | 56 | 59 | 61 | 64 | 67 | 70 | 73 | 76 | 79 | 82 | 85 | 89 | 92 | 95 | 98 | 101 | 104 | 108 | 111 |
| | 0.05 | 57 | 60 | 63 | 67 | 70 | 73 | 76 | 80 | 83 | 87 | 90 | 93 | 97 | 100 | 104 | 107 | 111 | 114 | 118 |
| | 0.1 | 59 | 62 | 66 | 69 | 73 | 77 | 80 | 84 | 88 | 92 | 95 | 99 | 103 | 107 | 110 | 114 | 118 | 122 | 126 |

Table T-17: Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum(WRS) Statistic
(Continued)

| <i>n</i> | w_q | <i>m</i> | | | | | | | | | | | | | | | | | | |
|-----------|--------------|----------|-----|-----|-----|-----|-----|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 11 | 0.001 | 66 | 66 | 67 | 69 | 71 | 73 | 75 | 77 | 79 | 82 | 84 | 87 | 89 | 91 | 94 | 96 | 99 | 101 | 104 |
| | 0.005 | 66 | 67 | 69 | 72 | 74 | 77 | 80 | 83 | 85 | 88 | 91 | 94 | 97 | 100 | 103 | 106 | 109 | 112 | 115 |
| | 0.01 | 66 | 68 | 71 | 74 | 76 | 79 | 82 | 85 | 89 | 92 | 95 | 98 | 101 | 104 | 108 | 111 | 114 | 117 | 120 |
| | 0.025 | 67 | 70 | 73 | 76 | 80 | 83 | 86 | 90 | 93 | 97 | 100 | 104 | 107 | 111 | 114 | 118 | 122 | 125 | 129 |
| | 0.05 | 68 | 72 | 75 | 79 | 83 | 86 | 90 | 94 | 98 | 101 | 105 | 109 | 113 | 117 | 121 | 124 | 128 | 132 | 136 |
| | 0.1 | 70 | 74 | 78 | 82 | 86 | 90 | 94 | 98 | 103 | 107 | 111 | 115 | 119 | 124 | 128 | 132 | 136 | 140 | 145 |
| 12 | 0.001 | 78 | 78 | 79 | 81 | 83 | 86 | 88 | 91 | 93 | 96 | 98 | 102 | 104 | 106 | 110 | 113 | 116 | 118 | 121 |
| | 0.005 | 78 | 80 | 82 | 85 | 88 | 91 | 94 | 97 | 100 | 103 | 106 | 110 | 113 | 116 | 120 | 123 | 126 | 130 | 133 |
| | 0.01 | 78 | 81 | 84 | 87 | 90 | 93 | 96 | 100 | 103 | 107 | 110 | 114 | 117 | 121 | 125 | 128 | 132 | 135 | 139 |
| | 0.025 | 80 | 83 | 86 | 90 | 93 | 97 | 101 | 105 | 108 | 112 | 116 | 120 | 124 | 128 | 132 | 136 | 140 | 144 | 148 |
| | 0.05 | 81 | 84 | 88 | 92 | 96 | 100 | 105 | 109 | 111 | 117 | 121 | 126 | 130 | 134 | 139 | 143 | 147 | 151 | 156 |
| | 0.1 | 83 | 87 | 91 | 96 | 100 | 105 | 109 | 114 | 118 | 123 | 128 | 132 | 137 | 142 | 146 | 151 | 156 | 160 | 165 |
| 13 | 0.001 | 91 | 91 | 93 | 95 | 97 | 100 | 103 | 106 | 109 | 112 | 115 | 118 | 121 | 124 | 127 | 130 | 134 | 137 | 140 |
| | 0.005 | 91 | 93 | 95 | 99 | 102 | 105 | 109 | 112 | 116 | 119 | 123 | 126 | 130 | 134 | 137 | 141 | 145 | 149 | 152 |
| | 0.01 | 92 | 94 | 97 | 101 | 104 | 108 | 112 | 115 | 119 | 123 | 127 | 131 | 135 | 139 | 143 | 147 | 151 | 155 | 159 |
| | 0.025 | 93 | 96 | 100 | 104 | 108 | 112 | 116 | 120 | 125 | 129 | 133 | 137 | 142 | 146 | 151 | 155 | 159 | 164 | 168 |
| | 0.05 | 94 | 98 | 102 | 107 | 111 | 116 | 120 | 125 | 129 | 134 | 139 | 143 | 148 | 153 | 157 | 162 | 167 | 172 | 176 |
| | 0.1 | 96 | 101 | 105 | 110 | 115 | 120 | 125 | 130 | 135 | 140 | 145 | 150 | 155 | 160 | 166 | 171 | 176 | 181 | 186 |
| 14 | 0.001 | 105 | 105 | 107 | 109 | 112 | 115 | 118 | 121 | 125 | 128 | 131 | 135 | 138 | 142 | 145 | 149 | 152 | 156 | 160 |
| | 0.005 | 105 | 107 | 110 | 113 | 117 | 121 | 124 | 128 | 132 | 136 | 140 | 144 | 148 | 152 | 156 | 160 | 164 | 169 | 173 |
| | 0.01 | 106 | 108 | 112 | 116 | 119 | 123 | 128 | 132 | 136 | 140 | 144 | 149 | 153 | 157 | 162 | 166 | 171 | 175 | 179 |
| | 0.025 | 107 | 111 | 115 | 119 | 123 | 128 | 132 | 137 | 142 | 146 | 151 | 156 | 161 | 165 | 170 | 175 | 180 | 184 | 189 |
| | 0.05 | 109 | 113 | 117 | 122 | 127 | 132 | 137 | 142 | 147 | 152 | 157 | 162 | 167 | 172 | 177 | 183 | 188 | 193 | 198 |
| | 0.1 | 110 | 116 | 121 | 126 | 131 | 137 | 142 | 147 | 153 | 158 | 164 | 169 | 175 | 180 | 186 | 191 | 197 | 203 | 208 |

Table T-17: Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum(WRS) Statistic
(Continued)

| n | w_q | m | | | | | | | | | | | | | | | | | | |
|-----------|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 15 | 0.001 | 120 | 120 | 122 | 125 | 128 | 133 | 135 | 138 | 142 | 145 | 149 | 153 | 157 | 161 | 164 | 168 | 172 | 176 | 180 |
| | 0.005 | 120 | 123 | 126 | 129 | 133 | 137 | 141 | 145 | 150 | 154 | 158 | 163 | 167 | 172 | 176 | 181 | 185 | 190 | 194 |
| | 0.01 | 121 | 124 | 128 | 132 | 136 | 140 | 145 | 149 | 154 | 158 | 163 | 168 | 172 | 177 | 182 | 187 | 191 | 196 | 201 |
| | 0.025 | 122 | 126 | 131 | 135 | 140 | 145 | 150 | 155 | 160 | 165 | 170 | 175 | 180 | 185 | 191 | 196 | 201 | 206 | 211 |
| | 0.05 | 124 | 128 | 133 | 139 | 144 | 149 | 154 | 160 | 165 | 171 | 176 | 182 | 187 | 193 | 198 | 204 | 209 | 215 | 221 |
| | 0.1 | 126 | 131 | 137 | 143 | 148 | 154 | 160 | 166 | 172 | 178 | 184 | 189 | 195 | 201 | 207 | 213 | 219 | 225 | 231 |
| 16 | 0.001 | 136 | 136 | 139 | 142 | 145 | 148 | 152 | 156 | 160 | 164 | 168 | 172 | 176 | 180 | 185 | 189 | 193 | 197 | 202 |
| | 0.005 | 136 | 139 | 142 | 146 | 150 | 155 | 159 | 164 | 168 | 173 | 178 | 182 | 187 | 192 | 197 | 202 | 207 | 211 | 216 |
| | 0.01 | 137 | 140 | 144 | 149 | 153 | 158 | 163 | 168 | 173 | 178 | 183 | 188 | 193 | 198 | 203 | 208 | 213 | 219 | 224 |
| | 0.025 | 138 | 143 | 148 | 152 | 158 | 163 | 168 | 174 | 179 | 184 | 190 | 196 | 201 | 207 | 212 | 218 | 223 | 229 | 235 |
| | 0.05 | 140 | 145 | 151 | 156 | 162 | 167 | 173 | 179 | 185 | 191 | 197 | 202 | 208 | 214 | 220 | 226 | 232 | 238 | 244 |
| | 0.1 | 142 | 148 | 154 | 160 | 166 | 173 | 179 | 185 | 191 | 198 | 204 | 211 | 217 | 223 | 230 | 236 | 243 | 249 | 256 |
| 17 | 0.001 | 153 | 154 | 156 | 159 | 163 | 167 | 171 | 175 | 179 | 183 | 188 | 192 | 197 | 201 | 206 | 211 | 215 | 220 | 224 |
| | 0.005 | 153 | 156 | 160 | 164 | 169 | 173 | 178 | 183 | 188 | 193 | 198 | 203 | 208 | 214 | 219 | 224 | 229 | 235 | 240 |
| | 0.01 | 154 | 158 | 162 | 167 | 172 | 177 | 182 | 187 | 192 | 198 | 203 | 209 | 214 | 220 | 225 | 231 | 236 | 242 | 247 |
| | 0.025 | 156 | 160 | 165 | 171 | 176 | 182 | 188 | 193 | 199 | 205 | 211 | 217 | 223 | 229 | 235 | 241 | 247 | 253 | 259 |
| | 0.05 | 157 | 163 | 169 | 174 | 180 | 187 | 193 | 199 | 205 | 211 | 218 | 224 | 231 | 237 | 243 | 250 | 256 | 263 | 269 |
| | 0.1 | 160 | 166 | 172 | 179 | 185 | 192 | 199 | 206 | 212 | 219 | 226 | 233 | 239 | 246 | 253 | 260 | 267 | 274 | 281 |
| 18 | 0.001 | 171 | 172 | 175 | 178 | 182 | 186 | 190 | 195 | 199 | 204 | 209 | 214 | 218 | 223 | 228 | 233 | 238 | 243 | 248 |
| | 0.005 | 171 | 174 | 178 | 183 | 188 | 193 | 198 | 203 | 209 | 214 | 219 | 225 | 230 | 236 | 242 | 247 | 253 | 259 | 264 |
| | 0.01 | 172 | 176 | 181 | 186 | 191 | 196 | 202 | 208 | 213 | 219 | 225 | 231 | 237 | 242 | 248 | 254 | 260 | 266 | 272 |
| | 0.025 | 174 | 179 | 184 | 190 | 196 | 202 | 208 | 214 | 220 | 227 | 233 | 239 | 246 | 252 | 258 | 265 | 271 | 278 | 284 |
| | 0.05 | 176 | 181 | 188 | 194 | 200 | 207 | 213 | 220 | 227 | 233 | 240 | 247 | 254 | 260 | 267 | 274 | 281 | 288 | 295 |
| | 0.1 | 178 | 185 | 192 | 199 | 206 | 213 | 220 | 227 | 234 | 241 | 249 | 256 | 263 | 270 | 278 | 285 | 292 | 300 | 307 |
| 19 | 0.001 | 190 | 191 | 194 | 198 | 202 | 206 | 211 | 216 | 220 | 225 | 231 | 236 | 241 | 246 | 251 | 257 | 262 | 268 | 273 |
| | 0.005 | 191 | 194 | 198 | 203 | 208 | 213 | 219 | 224 | 230 | 236 | 242 | 248 | 254 | 260 | 265 | 272 | 278 | 284 | 290 |
| | 0.01 | 192 | 195 | 200 | 206 | 211 | 217 | 223 | 229 | 235 | 241 | 247 | 254 | 260 | 266 | 273 | 279 | 285 | 292 | 298 |
| | 0.025 | 193 | 198 | 204 | 210 | 216 | 223 | 229 | 236 | 243 | 249 | 256 | 263 | 269 | 276 | 283 | 290 | 297 | 304 | 310 |
| | 0.05 | 195 | 201 | 208 | 214 | 221 | 228 | 235 | 242 | 249 | 256 | 263 | 271 | 278 | 285 | 292 | 300 | 307 | 314 | 321 |
| | 0.1 | 198 | 205 | 212 | 219 | 227 | 234 | 242 | 249 | 257 | 264 | 272 | 280 | 288 | 295 | 303 | 311 | 319 | 326 | 334 |

Table T-17: Quantiles, $w_q(n, m)$, of the Wilcoxon rank sum(WRS) Statistic
(Continued)

| n | w_q | m | | | | | | | | | | | | | | | | | | |
|-----------|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 20 | 0.001 | 210 | 211 | 214 | 218 | 223 | 227 | 232 | 237 | 243 | 248 | 253 | 259 | 265 | 270 | 276 | 281 | 287 | 293 | 299 |
| | 0.005 | 211 | 214 | 219 | 224 | 229 | 235 | 241 | 247 | 253 | 259 | 265 | 271 | 278 | 284 | 290 | 297 | 303 | 310 | 316 |
| | 0.01 | 212 | 216 | 221 | 227 | 233 | 239 | 245 | 251 | 258 | 264 | 271 | 278 | 284 | 291 | 298 | 304 | 311 | 318 | 325 |
| | 0.025 | 213 | 219 | 225 | 231 | 238 | 245 | 251 | 259 | 266 | 273 | 280 | 287 | 294 | 301 | 309 | 316 | 323 | 330 | 338 |
| | 0.05 | 215 | 222 | 229 | 236 | 243 | 250 | 258 | 265 | 273 | 280 | 288 | 295 | 303 | 311 | 318 | 326 | 334 | 341 | 349 |
| | 0.1 | 218 | 226 | 233 | 241 | 249 | 257 | 265 | 273 | 281 | 289 | 297 | 305 | 313 | 321 | 330 | 338 | 346 | 354 | 362 |

Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © 1980, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.,

Table T-18: Quantiles of the Squared Ranks statistic (see Section 25.13)

| <i>n</i> | <i>p</i> | <i>m</i> | | | | | | | |
|----------|----------|----------|-----|-----|-----|-----|-----|------|------|
| | | 3 | 4 | 5 | 6 | 7 | 9 | 9 | 10 |
| 3 | 0.005 | 14 | 14 | 14 | 14 | 14 | 14 | 21 | 21 |
| | 0.01 | 14 | 14 | 14 | 14 | 21 | 21 | 26 | 26 |
| | 0.025 | 14 | 14 | 21 | 26 | 29 | 30 | 35 | 41 |
| | 0.05 | 21 | 21 | 26 | 30 | 38 | 42 | 49 | 54 |
| | 0.1 | 26 | 29 | 35 | 42 | 50 | 59 | 69 | 77 |
| | 0.9 | 65 | 90 | 117 | 149 | 182 | 221 | 260 | 305 |
| | 0.95 | 70 | 101 | 129 | 161 | 197 | 238 | 285 | 333 |
| | 0.975 | 77 | 110 | 138 | 170 | 213 | 257 | 308 | 362 |
| | 0.99 | 77 | 110 | 149 | 194 | 230 | 285 | 329 | 394 |
| | 0.995 | 77 | 110 | 149 | 194 | 245 | 302 | 346 | 413 |
| 4 | 0.005 | 30 | 30 | 30 | 39 | 39 | 46 | 50 | 54 |
| | 0.01 | 30 | 30 | 39 | 46 | 50 | 51 | 62 | 66 |
| | 0.025 | 30 | 39 | 50 | 54 | 63 | 71 | 78 | 90 |
| | 0.05 | 39 | 50 | 57 | 66 | 78 | 90 | 102 | 114 |
| | 0.1 | 50 | 62 | 71 | 85 | 99 | 114 | 130 | 149 |
| | 0.9 | 111 | 142 | 182 | 222 | 270 | 321 | 375 | 435 |
| | 0.95 | 119 | 154 | 197 | 246 | 294 | 350 | 413 | 476 |
| | 0.975 | 126 | 165 | 206 | 255 | 311 | 374 | 439 | 510 |
| | 0.99 | 126 | 174 | 219 | 270 | 334 | 401 | 470 | 545 |
| | 0.995 | 126 | 174 | 230 | 281 | 351 | 414 | 494 | 567 |
| 5 | 0.005 | 55 | 55 | 66 | 75 | 79 | 88 | 99 | 110 |
| | 0.01 | 55 | 66 | 75 | 82 | 90 | 103 | 115 | 127 |
| | 0.025 | 66 | 79 | 88 | 100 | 114 | 130 | 145 | 162 |
| | 0.05 | 75 | 88 | 103 | 120 | 135 | 155 | 175 | 195 |
| | 0.1 | 87 | 103 | 121 | 142 | 163 | 187 | 212 | 239 |
| | 0.9 | 169 | 214 | 264 | 319 | 379 | 445 | 514 | 591 |
| | 0.95 | 178 | 228 | 282 | 342 | 410 | 479 | 558 | 639 |
| | 0.975 | 183 | 235 | 297 | 363 | 433 | 508 | 592 | 680 |
| | 0.99 | 190 | 246 | 310 | 382 | 459 | 543 | 631 | 727 |
| | 0.995 | 190 | 255 | 319 | 391 | 478 | 559 | 654 | 754 |
| 6 | 0.005 | 91 | 104 | 115 | 124 | 136 | 152 | 167 | 182 |
| | 0.01 | 91 | 115 | 124 | 139 | 155 | 175 | 191 | 210 |
| | 0.025 | 115 | 130 | 143 | 164 | 184 | 208 | 231 | 255 |
| | 0.05 | 124 | 139 | 164 | 187 | 211 | 239 | 268 | 299 |
| | 0.1 | 136 | 163 | 187 | 215 | 247 | 280 | 315 | 352 |
| | 0.9 | 243 | 300 | 364 | 435 | 511 | 592 | 679 | 772 |
| | 0.95 | 255 | 319 | 386 | 463 | 545 | 634 | 730 | 831 |
| | 0.975 | 259 | 331 | 406 | 486 | 574 | 670 | 771 | 880 |
| | 0.99 | 271 | 339 | 424 | 511 | 607 | 706 | 817 | 935 |
| | 0.995 | 271 | 346 | 431 | 526 | 624 | 731 | 847 | 970 |
| 7 | 0.005 | 140 | 155 | 172 | 195 | 212 | 235 | 257 | 280 |
| | 0.01 | 155 | 172 | 191 | 212 | 236 | 260 | 287 | 315 |
| | 0.025 | 172 | 195 | 217 | 245 | 274 | 305 | 338 | 372 |
| | 0.05 | 188 | 212 | 240 | 274 | 308 | 344 | 384 | 425 |
| | 0.1 | 203 | 236 | 271 | 308 | 350 | 394 | 440 | 489 |
| | 0.9 | 335 | 407 | 487 | 572 | 665 | 764 | 871 | 984 |
| | 0.95 | 347 | 428 | 515 | 608 | 707 | 814 | 929 | 1051 |
| | 0.975 | 356 | 443 | 536 | 635 | 741 | 856 | 979 | 1108 |
| | 0.99 | 364 | 456 | 560 | 664 | 779 | 900 | 1032 | 1172 |
| | 0.995 | 371 | 467 | 571 | 683 | 803 | 929 | 1067 | 1212 |

Table T-18: Quantiles of the Squared Ranks statistic (Continued)

| <i>n</i> | <i>p</i> | <i>m</i> | | | | | | | |
|-----------|--------------|----------|-----|------|------|------|------|------|------|
| | | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 8 | 0.005 | 204 | 236 | 260 | 284 | 311 | 340 | 368 | 401 |
| | 0.01 | 221 | 249 | 276 | 309 | 340 | 372 | 408 | 445 |
| | 0.025 | 249 | 276 | 311 | 345 | 384 | 425 | 468 | 513 |
| | 0.05 | 268 | 300 | 340 | 381 | 426 | 473 | 524 | 576 |
| | 0.1 | 285 | 329 | 374 | 423 | 476 | 531 | 590 | 652 |
| | 0.9 | 447 | 536 | 632 | 735 | 846 | 965 | 1091 | 1224 |
| | 0.95 | 464 | 560 | 664 | 776 | 896 | 1023 | 1159 | 1303 |
| | 0.975 | 476 | 579 | 689 | 807 | 935 | 1071 | 1215 | 1368 |
| | 0.99 | 485 | 599 | 716 | 840 | 980 | 1124 | 1277 | 1442 |
| | 0.995 | 492 | 604 | 731 | 863 | 1005 | 1156 | 1319 | 1489 |
| 9 | 0.005 | 304 | 325 | 361 | 393 | 429 | 466 | 508 | 549 |
| | 0.01 | 321 | 349 | 384 | 423 | 464 | 508 | 553 | 601 |
| | 0.025 | 342 | 380 | 423 | 469 | 517 | 570 | 624 | 682 |
| | 0.05 | 365 | 406 | 457 | 510 | 567 | 626 | 689 | 755 |
| | 0.1 | 390 | 444 | 501 | 561 | 625 | 694 | 766 | 843 |
| | 0.9 | 581 | 689 | 803 | 925 | 1056 | 1195 | 1343 | 1498 |
| | 0.95 | 601 | 717 | 840 | 972 | 1112 | 1251 | 1420 | 1587 |
| | 0.975 | 615 | 741 | 870 | 1009 | 1158 | 1317 | 1485 | 1662 |
| | 0.99 | 624 | 757 | 900 | 1049 | 1209 | 1377 | 1556 | 1745 |
| | 0.995 | 629 | 769 | 916 | 1073 | 1239 | 1417 | 1601 | 1798 |
| 10 | 0.005 | 406 | 448 | 486 | 526 | 573 | 620 | 672 | 725 |
| | 0.01 | 425 | 470 | 513 | 561 | 613 | 667 | 725 | 785 |
| | 0.025 | 457 | 505 | 560 | 616 | 677 | 741 | 808 | 879 |
| | 0.05 | 486 | 539 | 601 | 665 | 734 | 806 | 883 | 963 |
| | 0.1 | 514 | 580 | 649 | 724 | 801 | 885 | 972 | 1064 |
| | 0.9 | 742 | 866 | 1001 | 1144 | 1296 | 1457 | 1627 | 1806 |
| | 0.95 | 765 | 901 | 1045 | 1197 | 1360 | 1533 | 1715 | 1907 |
| | 0.975 | 778 | 925 | 1078 | 1241 | 1413 | 1596 | 1788 | 1991 |
| | 0.99 | 793 | 949 | 1113 | 1286 | 1470 | 1664 | 1869 | 2085 |
| | 0.995 | 798 | 961 | 1130 | 1314 | 1505 | 1708 | 1921 | 2145 |

Conover, W. J., 1980, Practical Nonparametric Statistics, 2nd Ed., © John Wiley & Sons, Inc. Adapted from tables by Conover and Iman, 1978, Communications in Statistics, B7. 491-513, Marcel-Dekker Journals, New York. Reprinted with the permission of John Wiley & Sons, Inc., Marcel- Dekker, Inc.

Table T-19: Quantiles, w_p , of the Spearman statistic, r_s , (see Section 25.15)

| n | q | | | | | |
|-----|--------|--------|--------|--------|--------|--------|
| | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 |
| 4 | 0.8000 | 0.8000 | | | | |
| 5 | 0.7000 | 0.8000 | 0.9000 | 0.9000 | | |
| 6 | 0.6000 | 0.7714 | 0.8286 | 0.8857 | 0.9429 | |
| 7 | 0.5357 | 0.6786 | 0.7450 | 0.8571 | 0.8929 | 0.9643 |
| 8 | 0.5000 | 0.6190 | 0.7143 | 0.8095 | 0.8571 | 0.9286 |
| 9 | 0.4667 | 0.5833 | 0.6833 | 0.7667 | 0.8167 | 0.9000 |
| 10 | 0.4424 | 0.5515 | 0.6364 | 0.7333 | 0.7818 | 0.8667 |
| 11 | 0.4182 | 0.5273 | 0.6091 | 0.7000 | 0.7455 | 0.8364 |
| 12 | 0.3986 | 0.4965 | 0.5804 | 0.6713 | 0.7273 | 0.8182 |
| 13 | 0.3791 | 0.4780 | 0.5549 | 0.6429 | 0.6978 | 0.7912 |
| 14 | 0.3626 | 0.4593 | 0.5341 | 0.6220 | 0.6747 | 0.7670 |
| 15 | 0.3500 | 0.4429 | 0.5179 | 0.6000 | 0.6536 | 0.7464 |
| 16 | 0.3382 | 0.4265 | 0.5000 | 0.5824 | 0.6324 | 0.7265 |
| 17 | 0.3260 | 0.4118 | 0.4853 | 0.5637 | 0.6152 | 0.7083 |
| 18 | 0.3148 | 0.3994 | 0.4716 | 0.5480 | 0.5975 | 0.6904 |
| 19 | 0.3070 | 0.3895 | 0.4579 | 0.5333 | 0.5825 | 0.6737 |
| 20 | 0.2977 | 0.3789 | 0.4451 | 0.5203 | 0.5684 | 0.6586 |
| 21 | 0.2909 | 0.3688 | 0.4351 | 0.5078 | 0.5545 | 0.6455 |
| 22 | 0.2829 | 0.3597 | 0.4241 | 0.4963 | 0.5426 | 0.6318 |
| 23 | 0.2767 | 0.3518 | 0.4150 | 0.4852 | 0.5306 | 0.6186 |
| 24 | 0.2704 | 0.3435 | 0.4061 | 0.4748 | 0.5200 | 0.6070 |
| 25 | 0.2646 | 0.3362 | 0.3977 | 0.4654 | 0.5100 | 0.5962 |
| 26 | 0.2588 | 0.3299 | 0.3894 | 0.4564 | 0.5002 | 0.5856 |
| 27 | 0.2540 | 0.3236 | 0.3822 | 0.4481 | 0.4915 | 0.5757 |
| 28 | 0.2490 | 0.3175 | 0.3749 | 0.4401 | 0.4828 | 0.5660 |
| 29 | 0.2443 | 0.3113 | 0.3685 | 0.4320 | 0.4744 | 0.5567 |
| 30 | 0.2400 | 0.3059 | 0.3620 | 0.4251 | 0.4665 | 0.5479 |

Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., © John Wiley & Sons, Inc. Adapted from G. J. Glasser and R. F. Winter, 1961, *Biometrika*, 49; 444. Reprinted with the permission of John Wiley & Sons, Inc., the *Biometrika* Trustees,

This table presents quantiles w_p for several values of p for testing the statistic R_s known as the Spearman rank correlation coefficient.

Quantiles of the lower end are computed by the equation $w_p = -w_{1-p}$

Lower tail critical regions are R_s greater than but not equal to the listed quantile.

Upper tail critical regions are R_s less than but not equal to the listed quantile.

Table T-20: Critical values for Grubbs' *T*-test (see Section 26.5)

| Number of Observations
<i>n</i> | Upper significant level, α | | | | | |
|------------------------------------|-----------------------------------|-------|-------|-------|-------|-------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 3 | 1.155 | 1.155 | 1.155 | 1.155 | 1.153 | 1.148 |
| 4 | 1.499 | 1.496 | 1.492 | 1.481 | 1.463 | 1.425 |
| 5 | 1.780 | 1.764 | 1.749 | 1.715 | 1.672 | 1.602 |
| 6 | 2.011 | 1.973 | 1.944 | 1.887 | 1.822 | 1.729 |
| 7 | 2.201 | 2.139 | 2.097 | 2.020 | 1.938 | 1.828 |
| 8 | 2.358 | 2.274 | 2.221 | 2.126 | 2.032 | 1.909 |
| 9 | 2.492 | 2.387 | 2.323 | 2.215 | 2.110 | 1.977 |
| 10 | 2.606 | 2.482 | 2.410 | 2.290 | 2.176 | 2.036 |
| 11 | 2.705 | 2.564 | 2.485 | 2.355 | 2.234 | 2.088 |
| 12 | 2.791 | 2.636 | 2.550 | 2.412 | 2.285 | 2.134 |
| 13 | 2.867 | 2.699 | 2.607 | 2.462 | 2.331 | 2.175 |
| 14 | 2.935 | 2.755 | 2.659 | 2.507 | 2.371 | 2.213 |
| 15 | 2.997 | 2.806 | 2.705 | 2.549 | 2.409 | 2.247 |
| 16 | 3.052 | 2.852 | 2.747 | 2.585 | 2.443 | 2.279 |
| 17 | 3.103 | 2.894 | 2.785 | 2.620 | 2.475 | 2.309 |
| 18 | 3.149 | 2.932 | 2.821 | 2.651 | 2.504 | 2.335 |
| 19 | 3.191 | 2.968 | 2.854 | 2.681 | 2.532 | 2.361 |
| 20 | 3.230 | 3.001 | 2.884 | 2.709 | 2.557 | 2.385 |
| 21 | 3.266 | 3.031 | 2.912 | 2.733 | 2.580 | 2.408 |
| 22 | 3.300 | 3.060 | 2.939 | 2.758 | 2.603 | 2.429 |
| 23 | 3.332 | 3.087 | 2.963 | 2.781 | 2.624 | 2.448 |
| 24 | 3.362 | 3.112 | 2.987 | 2.802 | 2.644 | 2.467 |
| 25 | 3.389 | 3.135 | 3.009 | 2.822 | 2.663 | 2.486 |
| 26 | 3.415 | 3.157 | 3.029 | 2.841 | 2.681 | 2.502 |
| 27 | 3.440 | 3.178 | 3.049 | 2.859 | 2.698 | 2.519 |
| 28 | 3.464 | 3.199 | 3.068 | 2.876 | 2.714 | 2.534 |
| 29 | 3.486 | 3.218 | 3.085 | 2.893 | 2.730 | 2.549 |
| 30 | 3.507 | 3.236 | 3.103 | 2.908 | 2.745 | 2.563 |
| 31 | 3.528 | 3.253 | 3.119 | 2.924 | 2.759 | 2.577 |
| 32 | 3.546 | 3.270 | 3.135 | 2.938 | 2.773 | 2.591 |
| 33 | 3.565 | 3.286 | 3.150 | 2.952 | 2.786 | 2.604 |
| 34 | 3.582 | 3.301 | 3.164 | 2.965 | 2.799 | 2.616 |
| 35 | 3.599 | 3.316 | 3.178 | 2.979 | 2.811 | 2.628 |
| 36 | 3.616 | 3.330 | 3.191 | 2.991 | 2.823 | 2.639 |
| 37 | 3.631 | 3.343 | 3.204 | 3.003 | 2.835 | 2.650 |
| 38 | 3.646 | 3.356 | 3.216 | 3.014 | 2.846 | 2.661 |
| 39 | 3.660 | 3.369 | 3.228 | 3.025 | 2.857 | 2.671 |
| 40 | 3.673 | 3.381 | 3.240 | 3.036 | 2.866 | 2.682 |
| 41 | 3.687 | 3.393 | 3.251 | 3.046 | 2.877 | 2.692 |
| 42 | 3.700 | 3.404 | 3.261 | 3.057 | 2.887 | 2.700 |

Table T-20: Critical values for Grubbs' T -test (Continued)

| Number of
Observation
n | Upper significant level, α | | | | | |
|---------------------------------|-----------------------------------|-------|-------|-------|-------|-------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 43 | 3.712 | 3.415 | 3.271 | 3.067 | 2.896 | 2.710 |
| 44 | 3.724 | 3.425 | 3.282 | 3.075 | 2.905 | 2.719 |
| 45 | 3.736 | 3.435 | 3.292 | 3.085 | 2.914 | 2.727 |
| 46 | 3.747 | 3.445 | 3.302 | 3.094 | 2.923 | 2.736 |
| 47 | 3.757 | 3.455 | 3.310 | 3.103 | 2.931 | 2.744 |
| 48 | 3.768 | 3.464 | 3.319 | 3.111 | 2.940 | 2.753 |
| 49 | 3.779 | 3.474 | 3.329 | 3.120 | 2.948 | 2.760 |
| 50 | 3.789 | 3.483 | 3.336 | 3.128 | 2.956 | 2.768 |
| 51 | 3.798 | 3.491 | 3.345 | 3.136 | 2.964 | 2.775 |
| 52 | 3.808 | 3.500 | 3.353 | 3.143 | 2.971 | 2.783 |
| 53 | 3.816 | 3.507 | 3.361 | 3.151 | 2.978 | 2.790 |
| 54 | 3.825 | 3.516 | 3.368 | 3.158 | 2.986 | 2.798 |
| 55 | 3.834 | 3.524 | 3.376 | 3.166 | 2.992 | 2.804 |
| 56 | 3.842 | 3.531 | 3.383 | 3.172 | 3.000 | 2.811 |
| 57 | 3.851 | 3.539 | 3.391 | 3.180 | 3.006 | 2.818 |
| 58 | 3.858 | 3.546 | 3.397 | 3.186 | 3.013 | 2.824 |
| 59 | 3.867 | 3.553 | 3.405 | 3.193 | 3.019 | 2.831 |
| 60 | 3.874 | 3.560 | 3.411 | 3.199 | 3.025 | 2.837 |
| 61 | 3.882 | 3.566 | 3.418 | 3.205 | 3.032 | 2.842 |
| 62 | 3.889 | 3.573 | 3.424 | 3.212 | 3.037 | 2.849 |
| 63 | 3.896 | 3.579 | 3.430 | 3.218 | 3.044 | 2.854 |
| 64 | 3.903 | 3.586 | 3.437 | 3.224 | 3.049 | 2.860 |
| 65 | 3.910 | 3.592 | 3.442 | 3.230 | 3.055 | 2.866 |
| 66 | 3.917 | 3.598 | 3.449 | 3.235 | 3.061 | 2.871 |
| 67 | 3.923 | 3.605 | 3.454 | 3.241 | 3.066 | 2.877 |
| 68 | 3.930 | 3.610 | 3.460 | 3.246 | 3.071 | 2.883 |
| 69 | 3.936 | 3.617 | 3.466 | 3.252 | 3.076 | 2.888 |
| 70 | 3.942 | 3.622 | 3.471 | 3.257 | 3.082 | 2.893 |
| 71 | 3.948 | 3.627 | 3.476 | 3.262 | 3.087 | 2.897 |
| 72 | 3.954 | 3.633 | 3.482 | 3.267 | 3.092 | 2.903 |
| 73 | 3.960 | 3.638 | 3.487 | 3.272 | 3.098 | 2.908 |
| 74 | 3.965 | 3.643 | 3.492 | 3.278 | 3.102 | 2.912 |
| 75 | 3.971 | 3.648 | 3.496 | 3.282 | 3.107 | 2.917 |

Table T-20: Critical values for Grubbs' *T*-test (Continued)

| Number of Observation
<i>n</i> | Upper significant level, α | | | | | |
|-----------------------------------|-----------------------------------|-------|-------|-------|-------|-------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 76 | 3.977 | 3.654 | 3.502 | 3.287 | 3.111 | 2.922 |
| 77 | 3.982 | 3.658 | 3.507 | 3.291 | 3.117 | 2.927 |
| 78 | 3.987 | 3.663 | 3.511 | 3.297 | 3.121 | 2.931 |
| 79 | 3.992 | 3.669 | 3.516 | 3.301 | 3.125 | 2.935 |
| 80 | 3.998 | 3.673 | 3.521 | 3.305 | 3.130 | 2.940 |
| 81 | 4.002 | 3.677 | 3.525 | 3.309 | 3.134 | 2.945 |
| 82 | 4.007 | 3.682 | 3.529 | 3.315 | 3.139 | 2.949 |
| 83 | 4.012 | 3.687 | 3.534 | 3.319 | 3.143 | 2.953 |
| 84 | 4.017 | 3.691 | 3.539 | 3.323 | 3.147 | 2.957 |
| 85 | 4.021 | 3.695 | 3.543 | 3.327 | 3.151 | 2.961 |
| 86 | 4.026 | 3.699 | 3.547 | 3.331 | 3.155 | 2.966 |
| 87 | 4.031 | 3.704 | 3.551 | 3.335 | 3.160 | 2.970 |
| 88 | 4.035 | 3.708 | 3.555 | 3.339 | 3.163 | 2.973 |
| 89 | 4.039 | 3.712 | 3.559 | 3.343 | 3.167 | 2.977 |
| 90 | 4.044 | 3.716 | 3.563 | 3.347 | 3.171 | 2.981 |
| 91 | 4.049 | 3.720 | 3.567 | 3.350 | 3.174 | 2.984 |
| 92 | 4.053 | 3.725 | 3.570 | 3.355 | 3.179 | 2.989 |
| 93 | 4.057 | 3.728 | 3.575 | 3.358 | 3.182 | 2.993 |
| 94 | 4.060 | 3.732 | 3.579 | 3.362 | 3.186 | 2.996 |
| 95 | 4.064 | 3.736 | 3.582 | 3.365 | 3.189 | 3.000 |
| 96 | 4.069 | 3.739 | 3.586 | 3.369 | 3.193 | 3.003 |
| 97 | 4.073 | 3.744 | 3.589 | 3.372 | 3.196 | 3.006 |
| 98 | 4.076 | 3.747 | 3.593 | 3.377 | 3.201 | 3.011 |
| 99 | 4.080 | 3.750 | 3.597 | 3.380 | 3.204 | 3.014 |
| 100 | 4.084 | 3.754 | 3.600 | 3.383 | 3.207 | 3.017 |
| 101 | 4.088 | 3.757 | 3.603 | 3.386 | 3.210 | 3.021 |
| 102 | 4.092 | 3.760 | 3.607 | 3.390 | 3.214 | 3.024 |
| 103 | 4.095 | 3.765 | 3.610 | 3.393 | 3.217 | 3.027 |
| 104 | 4.098 | 3.768 | 3.614 | 3.397 | 3.220 | 3.030 |
| 105 | 4.102 | 3.771 | 3.617 | 3.400 | 3.224 | 3.033 |
| 106 | 4.105 | 3.774 | 3.620 | 3.403 | 3.227 | 3.037 |
| 107 | 4.109 | 3.777 | 3.623 | 3.406 | 3.230 | 3.040 |
| 108 | 4.112 | 3.780 | 3.626 | 3.409 | 3.233 | 3.043 |
| 109 | 4.116 | 3.784 | 3.629 | 3.412 | 3.236 | 3.046 |
| 110 | 4.119 | 3.787 | 3.632 | 3.415 | 3.239 | 3.049 |
| 111 | 4.122 | 3.790 | 3.636 | 3.418 | 3.242 | 3.052 |
| 112 | 4.125 | 3.793 | 3.639 | 3.422 | 3.245 | 3.055 |

Table T-20: Critical values for Grubbs' T -test (Continued)

| Number of
Observations,
n | Upper significant level, α | | | | | |
|-----------------------------------|-----------------------------------|-------|-------|-------|-------|-------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 113 | 4.129 | 3.796 | 3.642 | 3.424 | 3.248 | 3.058 |
| 114 | 4.132 | 3.799 | 3.645 | 3.427 | 3.251 | 3.061 |
| 115 | 4.135 | 3.802 | 3.647 | 3.430 | 3.254 | 3.064 |
| 116 | 4.138 | 3.805 | 3.650 | 3.433 | 3.257 | 3.067 |
| 117 | 4.141 | 3.808 | 3.653 | 3.435 | 3.259 | 3.070 |
| 118 | 4.144 | 3.811 | 3.656 | 3.438 | 3.262 | 3.073 |
| 119 | 4.146 | 3.814 | 3.659 | 3.441 | 3.265 | 3.075 |
| 120 | 4.150 | 3.817 | 3.662 | 3.444 | 3.267 | 3.078 |
| 121 | 4.153 | 3.819 | 3.665 | 3.447 | 3.270 | 3.081 |
| 122 | 4.156 | 3.822 | 3.667 | 3.450 | 3.274 | 3.083 |
| 123 | 4.159 | 3.824 | 3.670 | 3.452 | 3.276 | 3.086 |
| 124 | 4.161 | 3.827 | 3.672 | 3.455 | 3.279 | 3.089 |
| 125 | 4.164 | 3.831 | 3.675 | 3.457 | 3.281 | 3.092 |
| 126 | 4.166 | 3.833 | 3.677 | 3.460 | 3.284 | 3.095 |
| 127 | 4.169 | 3.836 | 3.680 | 3.462 | 3.286 | 3.097 |
| 128 | 4.173 | 3.838 | 3.683 | 3.465 | 3.289 | 3.100 |
| 129 | 4.175 | 3.840 | 3.686 | 3.467 | 3.291 | 3.102 |
| 130 | 4.178 | 3.843 | 3.688 | 3.470 | 3.294 | 3.104 |
| 131 | 4.180 | 3.845 | 3.690 | 3.473 | 3.296 | 3.107 |
| 132 | 4.183 | 3.848 | 3.693 | 3.475 | 3.298 | 3.109 |
| 133 | 4.185 | 3.850 | 3.695 | 3.478 | 3.302 | 3.112 |
| 134 | 4.188 | 3.853 | 3.697 | 3.480 | 3.304 | 3.114 |
| 135 | 4.190 | 3.856 | 3.700 | 3.482 | 3.306 | 3.116 |
| 136 | 4.193 | 3.858 | 3.702 | 3.484 | 3.309 | 3.119 |
| 137 | 4.196 | 3.860 | 3.704 | 3.487 | 3.311 | 3.122 |
| 138 | 4.198 | 3.863 | 3.707 | 3.489 | 3.313 | 3.124 |
| 139 | 4.200 | 3.865 | 3.710 | 3.491 | 3.315 | 3.126 |
| 140 | 4.203 | 3.867 | 3.712 | 3.493 | 3.318 | 3.129 |

Grubbs, F. E., *Technometrics*, TCMTA, Vol 11, No. 4, February 1969, pp. 1–21. Grubbs, F. E., and Beck, G *Technometrics*, TCMTA, Vol 14, No. 4, November 1972, pp. 847–854, with permission.

Table T-21: Critical values (One-Sided Test) for w/s (ratio of range to sample standard deviation) (see Section 26.5)

| Number of Observations,
<i>n</i> | Level of significance | | |
|-------------------------------------|-----------------------|-----------------|------------------|
| | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.001$ |
| 3 | 2.00 | 2.00 | 2.00 |
| 4 | 2.43 | 2.44 | 2.45 |
| 5 | 2.75 | 2.80 | 2.81 |
| 6 | 3.01 | 3.10 | 3.12 |
| 7 | 3.22 | 3.34 | 3.37 |
| 8 | 3.40 | 3.54 | 3.58 |
| 9 | 3.55 | 3.72 | 3.77 |
| 10 | 3.68 | 3.88 | 3.94 |
| 11 | 3.80 | 4.01 | 4.08 |
| 12 | 3.91 | 4.13 | 4.21 |
| 13 | 4.00 | 4.24 | 4.32 |
| 14 | 4.09 | 4.34 | 4.43 |
| 15 | 4.17 | 4.43 | 4.53 |
| 16 | 4.24 | 4.51 | 4.62 |
| 17 | 4.31 | 4.59 | 4.69 |
| 18 | 4.38 | 4.66 | 4.77 |
| 19 | 4.43 | 4.73 | 4.84 |
| 20 | 4.49 | 4.79 | 4.91 |
| 30 | 4.89 | 5.25 | 5.39 |
| 40 | 5.15 | 5.54 | 5.69 |
| 50 | 5.35 | 5.77 | 5.91 |
| 60 | 5.50 | 5.93 | 6.09 |
| 80 | 5.73 | 6.18 | 6.35 |
| 100 | 5.90 | 6.36 | 6.54 |
| 150 | 6.18 | 6.64 | 6.84 |
| 200 | 6.38 | 6.85 | 7.03 |

Grubbs, F. E., "Sample Criteria for Testing Outlying Observations," *Annals of Mathematical Statistics*, AASTA Vol 21, March 1950, pp. 27-58.

Grubbs, F. E., "Procedures for Detecting Outlying Observations in Samples," *Technometrics*, TCMTA, Vol 11, No. 4, February 1969, pp. 1-21, with permission.

Table T22: Critical values for $S_{2n-1,n}/S_2$, or $S_{21,2}/S_2$ for simultaneously testing the two largest or two smallest observations (see Section 26.5)

| Number of Observations
<i>n</i> | Lower significant level, α | | | | | |
|------------------------------------|-----------------------------------|--------|--------|--------|--------|--------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 4 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0008 | 0.0031 |
| 5 | 0.0003 | 0.0018 | 0.0035 | 0.0090 | 0.0183 | 0.0376 |
| 6 | 0.0039 | 0.0116 | 0.0186 | 0.0349 | 0.0564 | 0.0920 |
| 7 | 0.0135 | 0.0308 | 0.0440 | 0.0708 | 0.1020 | 0.1479 |
| 8 | 0.0290 | 0.0563 | 0.0750 | 0.1101 | 0.1478 | 0.1994 |
| 9 | 0.0489 | 0.0851 | 0.1082 | 0.1492 | 0.1909 | 0.2454 |
| 10 | 0.0714 | 0.1150 | 0.1414 | 0.1864 | 0.2305 | 0.2863 |
| 11 | 0.0953 | 0.1448 | 0.1736 | 0.2213 | 0.2667 | 0.3227 |
| 12 | 0.1198 | 0.1738 | 0.2043 | 0.2537 | 0.2996 | 0.3552 |
| 13 | 0.1441 | 0.2016 | 0.2333 | 0.2836 | 0.3295 | 0.3843 |
| 14 | 0.1680 | 0.2280 | 0.2605 | 0.3112 | 0.3568 | 0.4106 |
| 15 | 0.1912 | 0.2530 | 0.2859 | 0.3367 | 0.3818 | 0.4345 |
| 16 | 0.2136 | 0.2767 | 0.3098 | 0.3603 | 0.4048 | 0.4562 |
| 17 | 0.2350 | 0.2990 | 0.3321 | 0.3822 | 0.4259 | 0.4761 |
| 18 | 0.2556 | 0.3200 | 0.3530 | 0.4025 | 0.4455 | 0.4944 |
| 19 | 0.2752 | 0.3398 | 0.3725 | 0.4214 | 0.4636 | 0.5113 |
| 20 | 0.2939 | 0.3585 | 0.3909 | 0.4391 | 0.4804 | 0.5270 |
| 21 | 0.3118 | 0.3761 | 0.4082 | 0.4556 | 0.4961 | 0.5415 |
| 22 | 0.3288 | 0.3927 | 0.4245 | 0.4711 | 0.5107 | 0.5550 |
| 23 | 0.3450 | 0.4085 | 0.4398 | 0.4857 | 0.5244 | 0.5677 |
| 24 | 0.3605 | 0.4234 | 0.4543 | 0.4994 | 0.5373 | 0.5795 |
| 25 | 0.3752 | 0.4376 | 0.4680 | 0.5123 | 0.5495 | 0.5906 |
| 26 | 0.3893 | 0.4510 | 0.4810 | 0.5245 | 0.5609 | 0.6011 |
| 27 | 0.4027 | 0.4638 | 0.4933 | 0.5360 | 0.5717 | 0.6110 |
| 28 | 0.4156 | 0.4759 | 0.5050 | 0.5470 | 0.5819 | 0.6203 |
| 29 | 0.4279 | 0.4875 | 0.5162 | 0.5574 | 0.5916 | 0.6292 |
| 30 | 0.4397 | 0.4985 | 0.5268 | 0.5672 | 0.6008 | 0.6375 |
| 31 | 0.4510 | 0.5091 | 0.5369 | 0.5766 | 0.6095 | 0.6455 |
| 32 | 0.4618 | 0.5192 | 0.5465 | 0.5856 | 0.6178 | 0.6530 |
| 33 | 0.4722 | 0.5288 | 0.5557 | 0.5941 | 0.6257 | 0.6602 |
| 34 | 0.4821 | 0.5381 | 0.5646 | 0.6023 | 0.6333 | 0.6671 |
| 35 | 0.4917 | 0.5469 | 0.5730 | 0.6101 | 0.6405 | 0.6737 |
| 36 | 0.5009 | 0.5554 | 0.5811 | 0.6175 | 0.6474 | 0.6800 |
| 37 | 0.5098 | 0.5636 | 0.5889 | 0.6247 | 0.6541 | 0.6860 |
| 38 | 0.5184 | 0.5714 | 0.5963 | 0.6316 | 0.6604 | 0.6917 |
| 39 | 0.5266 | 0.5789 | 0.6035 | 0.6382 | 0.6665 | 0.6972 |
| 40 | 0.5345 | 0.5862 | 0.6104 | 0.6445 | 0.6724 | 0.7025 |

Table T22: Critical values for $S_{2n-1,n}/S_2$, or $S_{21,2}/S_2$ for simultaneously testing the two largest or two smallest observations (Continued)

| Number of Observations
<i>n</i> | Lower significant level, α | | | | | |
|------------------------------------|-----------------------------------|--------|--------|--------|--------|--------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 79 | 0.7145 | 0.7477 | 0.7628 | 0.7836 | 0.8002 | 0.8180 |
| 80 | 0.7172 | 0.7501 | 0.7650 | 0.7856 | 0.8021 | 0.8197 |
| 81 | 0.7199 | 0.7525 | 0.7672 | 0.7876 | 0.8040 | 0.8213 |
| 82 | 0.7225 | 0.7548 | 0.7694 | 0.7896 | 0.8058 | 0.8230 |
| 83 | 0.7250 | 0.7570 | 0.7715 | 0.7915 | 0.8075 | 0.8245 |
| 84 | 0.7275 | 0.7592 | 0.7736 | 0.7934 | 0.8093 | 0.8261 |
| 85 | 0.7300 | 0.7614 | 0.7756 | 0.7953 | 0.8109 | 0.8276 |
| 86 | 0.7324 | 0.7635 | 0.7776 | 0.7971 | 0.8126 | 0.8291 |
| 87 | 0.7348 | 0.7656 | 0.7796 | 0.7989 | 0.8142 | 0.8306 |
| 88 | 0.7371 | 0.7677 | 0.7815 | 0.8006 | 0.8158 | 0.8321 |
| 89 | 0.7394 | 0.7697 | 0.7834 | 0.8023 | 0.8174 | 0.8335 |
| 90 | 0.7416 | 0.7717 | 0.7853 | 0.8040 | 0.8190 | 0.8349 |
| 91 | 0.7438 | 0.7736 | 0.7871 | 0.8057 | 0.8205 | 0.8362 |
| 92 | 0.7459 | 0.7755 | 0.7889 | 0.8073 | 0.8220 | 0.8376 |
| 93 | 0.7481 | 0.7774 | 0.7906 | 0.8089 | 0.8234 | 0.8389 |
| 94 | 0.7501 | 0.7792 | 0.7923 | 0.8104 | 0.8248 | 0.8402 |
| 95 | 0.7522 | 0.7810 | 0.7940 | 0.8120 | 0.8263 | 0.8414 |
| 96 | 0.7542 | 0.7828 | 0.7957 | 0.8135 | 0.8276 | 0.8427 |
| 97 | 0.7562 | 0.7845 | 0.7973 | 0.8149 | 0.8290 | 0.8439 |
| 98 | 0.7581 | 0.7862 | 0.7989 | 0.8164 | 0.8303 | 0.8451 |
| 99 | 0.7600 | 0.7879 | 0.8005 | 0.8178 | 0.8316 | 0.8463 |
| 100 | 0.7619 | 0.7896 | 0.8020 | 0.8192 | 0.8329 | 0.8475 |
| 101 | 0.7637 | 0.7912 | 0.8036 | 0.8206 | 0.8342 | 0.8486 |
| 102 | 0.7655 | 0.7928 | 0.8051 | 0.8220 | 0.8354 | 0.8497 |
| 103 | 0.7673 | 0.7944 | 0.8065 | 0.8233 | 0.8367 | 0.8508 |
| 104 | 0.7691 | 0.7959 | 0.8080 | 0.8246 | 0.8379 | 0.8519 |
| 105 | 0.7708 | 0.7974 | 0.8094 | 0.8259 | 0.8391 | 0.8530 |
| 106 | 0.7725 | 0.7989 | 0.8108 | 0.8272 | 0.8402 | 0.8541 |
| 107 | 0.7742 | 0.8004 | 0.8122 | 0.8284 | 0.8414 | 0.8551 |
| 108 | 0.7758 | 0.8018 | 0.8136 | 0.8297 | 0.8425 | 0.8563 |
| 109 | 0.7774 | 0.8033 | 0.8149 | 0.8309 | 0.8436 | 0.8571 |
| 110 | 0.7790 | 0.8047 | 0.8162 | 0.8321 | 0.8447 | 0.8581 |
| 111 | 0.7806 | 0.8061 | 0.8175 | 0.8333 | 0.8458 | 0.8591 |
| 112 | 0.7821 | 0.8074 | 0.8188 | 0.8344 | 0.8469 | 0.8600 |
| 113 | 0.7837 | 0.8088 | 0.8200 | 0.8356 | 0.8479 | 0.8610 |
| 114 | 0.7852 | 0.8101 | 0.8213 | 0.8367 | 0.8489 | 0.8619 |
| 115 | 0.7866 | 0.8114 | 0.8225 | 0.8378 | 0.8500 | 0.8628 |
| 116 | 0.7881 | 0.8127 | 0.8237 | 0.8389 | 0.8510 | 0.8637 |

Table T22: Critical values for $S_{2n-1,n}/S_2$, or $S_{21,2}/S_2$ for simultaneously testing the two largest or two smallest observations (Continued)

| Number of Observations
<i>n</i> | Lower significant level, α | | | | | |
|------------------------------------|-----------------------------------|--------|--------|--------|--------|--------|
| | 0.001 | 0.005 | 0.01 | 0.025 | 0.050 | 0.10 |
| 117 | 0.7895 | 0.8139 | 0.8249 | 0.8400 | 0.8519 | 0.8646 |
| 118 | 0.7909 | 0.8152 | 0.8261 | 0.8410 | 0.8529 | 0.8655 |
| 119 | 0.7923 | 0.8164 | 0.8272 | 0.8421 | 0.8539 | 0.8664 |
| 120 | 0.7937 | 0.8176 | 0.8284 | 0.8431 | 0.8548 | 0.8672 |
| 121 | 0.7951 | 0.8188 | 0.8295 | 0.8441 | 0.8557 | 0.8681 |
| 122 | 0.7964 | 0.8200 | 0.8306 | 0.8451 | 0.8567 | 0.8689 |
| 123 | 0.7977 | 0.8211 | 0.8317 | 0.8461 | 0.8576 | 0.8697 |
| 124 | 0.7990 | 0.8223 | 0.8327 | 0.8471 | 0.8585 | 0.8705 |
| 125 | 0.8003 | 0.8234 | 0.8338 | 0.8480 | 0.8593 | 0.8713 |
| 126 | 0.8016 | 0.8245 | 0.8348 | 0.8490 | 0.8602 | 0.8721 |
| 127 | 0.8028 | 0.8256 | 0.8359 | 0.8499 | 0.8611 | 0.8729 |
| 128 | 0.8041 | 0.8267 | 0.8369 | 0.8508 | 0.8619 | 0.8737 |
| 129 | 0.8053 | 0.8278 | 0.8379 | 0.8517 | 0.8627 | 0.8744 |
| 130 | 0.8065 | 0.8288 | 0.8389 | 0.8526 | 0.8636 | 0.8752 |
| 131 | 0.8077 | 0.8299 | 0.8398 | 0.8535 | 0.8644 | 0.8759 |
| 132 | 0.8088 | 0.8309 | 0.8408 | 0.8544 | 0.8652 | 0.8766 |
| 133 | 0.8100 | 0.8319 | 0.8418 | 0.8553 | 0.8660 | 0.8773 |
| 134 | 0.8111 | 0.8329 | 0.8427 | 0.8561 | 0.8668 | 0.8780 |
| 135 | 0.8122 | 0.8339 | 0.8436 | 0.8570 | 0.8675 | 0.8787 |
| 136 | 0.8134 | 0.8349 | 0.8445 | 0.8578 | 0.8683 | 0.8794 |
| 137 | 0.8145 | 0.8358 | 0.8454 | 0.8586 | 0.8690 | 0.8801 |
| 138 | 0.8155 | 0.8368 | 0.8463 | 0.8594 | 0.8698 | 0.8808 |
| 139 | 0.8166 | 0.8377 | 0.8472 | 0.8602 | 0.8705 | 0.8814 |
| 140 | 0.8176 | 0.8387 | 0.8481 | 0.8610 | 0.8712 | 0.8821 |

An observed ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

These significance levels are taken from Table 11, Grubbs, F. E., and Beck, G., "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations," *Technometrics*, TCMTA, Vol 14, No. 4, November 1972, pp. 847–854, with permission.

Table T-23: Dixon's criteria for testing for extreme values
(see Section 26.4)

| Statistic | <i>n</i> | $P_{0.70}$ | $P_{0.80}$ | $P_{0.90}$ | $P_{0.95}$ | $P_{0.98}$ | $P_{0.99}$ | $P_{0.995}$ |
|-----------|----------|------------|------------|------------|------------|------------|------------|-------------|
| r_{10} | 3 | 0.684 | 0.781 | 0.886 | 0.941 | 0.976 | 0.988 | 0.994 |
| | 4 | 0.471 | 0.560 | 0.679 | 0.765 | 0.846 | 0.889 | 0.926 |
| | 5 | 0.373 | 0.451 | 0.557 | 0.642 | 0.729 | 0.780 | 0.821 |
| | 6 | 0.318 | 0.386 | 0.482 | 0.560 | 0.644 | 0.698 | 0.740 |
| | 7 | 0.281 | 0.344 | 0.434 | 0.507 | 0.586 | 0.637 | 0.680 |
| r_{11} | 8 | 0.318 | 0.385 | 0.479 | 0.554 | 0.631 | 0.683 | 0.725 |
| | 9 | 0.288 | 0.352 | 0.441 | 0.512 | 0.587 | 0.635 | 0.677 |
| | 10 | 0.265 | 0.325 | 0.409 | 0.477 | 0.551 | 0.597 | 0.639 |
| r_{21} | 11 | 0.391 | 0.442 | 0.517 | 0.576 | 0.638 | 0.679 | 0.713 |
| | 12 | 0.370 | 0.419 | 0.490 | 0.546 | 0.605 | 0.642 | 0.675 |
| | 13 | 0.351 | 0.399 | 0.467 | 0.521 | 0.578 | 0.615 | 0.649 |
| r_{22} | 14 | 0.370 | 0.421 | 0.492 | 0.546 | 0.602 | 0.641 | 0.674 |
| | 15 | 0.353 | 0.402 | 0.472 | 0.525 | 0.579 | 0.616 | 0.647 |
| | 16 | 0.338 | 0.386 | 0.454 | 0.507 | 0.559 | 0.595 | 0.624 |
| | 17 | 0.325 | 0.373 | 0.438 | 0.490 | 0.542 | 0.577 | 0.605 |
| | 18 | 0.314 | 0.361 | 0.424 | 0.475 | 0.527 | 0.561 | 0.589 |
| | 19 | 0.304 | 0.350 | 0.412 | 0.462 | 0.514 | 0.547 | 0.575 |
| | 20 | 0.295 | 0.340 | 0.401 | 0.450 | 0.502 | 0.535 | 0.562 |
| | 21 | 0.287 | 0.331 | 0.391 | 0.440 | 0.491 | 0.524 | 0.551 |
| | 22 | 0.280 | 0.323 | 0.382 | 0.430 | 0.481 | 0.514 | 0.541 |
| | 23 | 0.274 | 0.316 | 0.374 | 0.421 | 0.472 | 0.505 | 0.532 |
| | 24 | 0.268 | 0.310 | 0.367 | 0.413 | 0.464 | 0.497 | 0.524 |
| | 25 | 0.262 | 0.304 | 0.360 | 0.406 | 0.457 | 0.489 | 0.516 |

Reproduced from Dixon (1953), with permission from the Biometric Society.

Bibliography

Note: These entries point to authors and material that we have found useful, insightful, and/or meaningful in our own pursuits of statistical quality and quantitative literacy. We believe that each of them, even if not mentioned specifically in the text, offers special rewards.

- Agresti, A. 1990, *Categorical Data Analysis*, John Wiley & Sons, Inc., New York, NY.
- Aitchison, J., and J. A. C. Brown, 1957, *The Lognormal Distribution with Special Reference to Its Uses in Econometrics*, University of Cambridge Department of Applied Economics Monograph: 5, Cambridge University Press.
- American National Standards Institute (ANSI) N15.15-1974, p. 12, *Assessment of the Assumption of Normality (Employing Individual Observed Values)*.
- Armitage, P., 1971, *Statistical Methods in Medical Research*, John Wiley & Sons, Inc., New York, NY.
- Babbie, E., 1990, *Survey Research Methods*, 2nd Ed., Wadsworth Publishing Company, Belmont, CA.
- Barclay, G.W., 1958, *Techniques of Population Analysis*, John Wiley & Sons, Inc., New York, NY.
- Barnett, V., 1991, *Sample Survey Principles and Methods*, Oxford University Press, New York, NY.
- Barnett, V., and T. Lewis, 1994, *Outliers in Statistical Data*, 3d Ed., John Wiley & Sons, Inc., New York, NY.
- Bartlett, M. S., 1937, Properties of Sufficiency and Statistical Tests, *Proceedings of the Royal Statistical Society*, Series A 160, pp. 268-282.
- Bates, D. M., and D. G. Watts, 1988, *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons, Inc., New York, NY.
- Bayes, T., 1763, An Essay Towards Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society*, **53**: 370-418.

- Beckman, R. J., and H. J. Trussell, 1974, The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression, *J. Am. Stat. Assoc.*, **69**: 199–201.
- Beckman, R. J., and R. D. Cook, 1983, Outlier.....s, *Technometrics*, **25**: 119-149.
- Beyer, W. H., 1974, *Handbook of Tables for Probability and Statistics*, 2nd Ed., The Chemical Rubber Co., Cleveland, OH.
- Bickel, P. J., E. A. Hammel, and J. W. O'Conner, 1975, Sex Bias in Graduate Admission: Data from Berkeley, *Science*, Vol. 187, No. 4175, pp. 398–404.
- Böhrer, A., 2008, One-sided and Two-sided Critical Values for Dixon's Outlier Test for Sample Sizes up to $n = 30$, *Economic quality Control*, **23**, No.1, 5 - 13.
- Bowen, W. M., and C. A. Bennett (Eds.), 1988, *Statistical Methods for Nuclear Material Management*, NUREG/CR-4604, U.S. Nuclear Regulatory Commission, U.S. Government Printing Office, Washington, DC.
- Bowker, A. H., and G. J. Lieberman, 1972, *Engineering Statistics*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Box, G. E. P., and N. R. Draper, 1987, *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, Inc., New York, NY.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter, 1978, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, Inc., New York, NY.
- Box, G. E. P. and G. M. Jenkins, 1976, *Time Series Analysis Forecast and Control*, Rev. Ed., Holden-Day, Inc., San Francisco, CA.
- Box G. E. P., and M. E. Muller, 1958, A Note of the Generation of Normal Deviates, *Annals of Mathematical Statistics*, **29**, 2
- Bradley, J. V., 1968, *Distribution-Free Statistical Tests*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Brownlee, K. A., 1965, *Statistical Theory and Methodology in Science and Engineering*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Burr, I. W., 1976, *Statistical Quality Control Methods*, Marcel Dekker, Inc., New York, NY.
- Burr, I. W., 1979, *Elementary Statistical Quality Control*, Marcel Dekker, Inc., New York, NY.

- Campbell, S. K., 1974, *Flaws and Fallacies in Statistical Thinking*, Prentice-Hall, Englewood Cliffs, NJ.
- Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey, 1983, *Graphical Methods for Data Analysis*, Duxbury Press, Boston, MA.
- Chapman, D. G., 1951, Some Properties of the Hypergeometric Distribution with Application to Zoological Sample Censuses, *University of California Publications in Statistics*, **1**: pp. 131–160, University of California Press, Berkeley, CA.
- Chatfield, C., 1991, Avoiding Statistical Pitfalls, *Statistical Science*, **6**: 240–268.
- Chernoff, H., and L. E. Moses, 1959, *Elementary Decision Theory*, John Wiley & Sons, Inc., New York, NY.
- Cleveland, W. S., 1985, *The Elements of Graphing Data*, Wadsworth Advanced Book Program, Monterey, CA.
- Cleveland, W. S., 1993, *Visualizing Data*, Hobart Press, Summit, NJ.
- Cleveland, W. S., and R. McGill, 1984, Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods, *Journal of the American Statistical Association*, **79**: 531–554.
- Cochran, W. G., 1952, The χ^2 Test of Goodness of Fit, *Annals of Mathematical Statistics*, **23**: 315–345.
- Cochran, W. G., 1954, Some Methods for Strengthening the Common χ^2 Tests, *Biometrics*, **10**: 417–451.
- Cochran, W. G., 1977, *Sampling Techniques*, 3rd Ed., John Wiley & Sons, Inc., New York, NY.
- Cochran, W. G., and G. M. Cox, 1957, *Experimental Design*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Coleman, D. E., 1985, Measuring Measurements, *RCA Engineer*, **30-3**: 16–23.
- Conover, W. J., 1980, *Practical Nonparametric Statistics*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Conover, W. J., and R. L. Iman, 1978, Some Exact Tables for the Squared Ranks Test, *Commun. Stat., Pt. B: Simul. Comput.*, **B7**: 491-513.
- Cramér, H., 1958, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.

- Crouse, C. F., 1968, *A Distribution-Free Method of Analyzing a 2^m Factorial Experiment*, *South African Statistics Journal*, **2**: 101408.
- D'Agostini, G., 1995, *Probability and Measurement Uncertainty in Physics—a Bayesian Primer*, Internal Report N. 1070 of the Dept. of Physics of the Rome University “La Sapienza”, and DESY-95-242.
- D'Agostino, R. B., 1971, An Omnibus Test of Normality for Moderate and Large Size Samples, *Biometrika*, Vol. 58.
- Dalal, R. H., E. B. Fowlkes, and B. Hoadley, 1989, Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure, *Journal of the American Statistical Association*, **84**: 945–957.
- Daniel, C., and F. S. Wood, with the assistance of J. W. Gorman, 1980, *Fitting Equations to Data: Computer Analysis of Multifactor Data*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Daniel, W. W., 1974, *Biostatistics: A Foundation for Analysis in the Health Sciences*, John Wiley & Sons, Inc., New York, NY.
- Davies, O. L. (Ed.), 1978, *The Design and Analysis of Industrial Experiments*, 2nd Ed., Hafner Publishing Co., New York, NY.
- Davies, O. L., and P. L. Goldsmith (Eds.), 1972, *Statistical Methods in Research and Production, with Special Reference to the Chemical Industry*, 4th Ed., Hafner Publishing Co., New York, NY.
- Deming, W. E., 1982, *Quality, Productivity, and Competitive Position*, MIT Center for Advanced Engineering Study, Cambridge, MA.
- Desu, M. and D. Raghavarao, 1990, *Sample Size Methodology*, Academic Press, Inc., San Diego, CA.
- Dixon, W. J., 1950, Analysis of Extreme Values, *Ann. Math. Stat.*, **21**: 488–506.
- Dixon, W. J. and F. J. Massey, 1983, *Introduction to Statistical Analysis*, 4th Ed., McGraw-Hill Book Company, Inc., New York, NY.
- Draper, N. R., and H. Smith, 1981, *Applied Regression Analysis*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Duncan, A. J., 1986, *Quality Control and Industrial Statistics*, 5th Ed., Irwin, Homewood, IL.
- Dunn, O. J., and V. A. Clark, 1987, *Applied Statistics: Analysis of Variance and Regression*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.

- Eadie, W. T., D. Dryard, F. E. James, M. Roos, and B. Sadoulet, 1971, *Statistical Methods in Experimental Physics*, American Elsevier Publishing Company, Inc., New York, NY.
- Eberstadt, N., 1995, *The Tyranny of Numbers: Mismeasurement and Misrule*, American Enterprise Institute for Public Policy Research, Washington, DC.
- Eisenhart, C., M. W. Hastay, and W. A. Wallis, 1947, *Techniques of Statistical Analysis*, McGraw-Hill Book Company, Inc., New York, NY.
- Evans, M., N. Hastings, B. Peacock, 2000, *Statistical Distributions*, 3rd Ed., John Wiley & Sons, Inc, New York, NY.
- Feller, W., 1957, *An Introduction to Probability Theory and Its Applications*, Volume I, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Ferguson, T. S., 1961, *Rules for Rejection of Outliers*, *Revue Inst. de Stat.* **3**: 29–43.
- Ferguson, T. S., 1967, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York, NY.
- Fisher, R. A., and E. A. Cornish, 1960, The Percentile Points of Distributions Having Known Cumulants, *Technometrics*, **2**: 209–225.
- Fisher, R. A., 1970, *Statistical Methods for Research Workers*, 14th Ed., Oliver and Boyd, Edinburgh, Scotland.
- Fleiss, J. L., 1981, *Statistical Methods for Rates and Proportions*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Frodesen, A. G., O. Skjeggstad, and H. Tofte, 1979, *Probability and Statistics in Particle Physics*, Universitetsforlaget, Bergen-Oslo-Tromsø, Norway.
- Ghiselli, E. E., 1964, *Theory of Psychological Measurement*, McGraw-Hill Book Company, Inc., New York, NY.
- Gibbons, J. D., 1976, *Nonparametric Methods for Quantitative Analysis*, Holt, Rinehart and Winston, , New York, NY.
- Gibbons, J. D., and S. Chakraborti, 1992, *Nonparametric Statistical Inference*, 3rd Ed., Marcel Dekker, Inc., New York, NY.
- Gilbert, N., 1976, *Statistics*, W. B. Saunders Co., Philadelphia, PA.

- Goldberger, A. S., 1964, *Econometric Theory*, John Wiley & Sons, Inc., New York, NY.
- Graybill, F. A., 1961, *An Introduction to Linear Statistical Models*, Volume I, McGraw-Hill Book Company, Inc., New York, NY.
- Grubbs, F. E., 1969, Procedures for Detecting Outlying Observations in Samples, *Technometrics*, **11**: 1–21.
- Guenther, W. C., 1964, *Analysis of Variance*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Guilford, J. P., 1954, *Psychometric Methods*, McGraw-Hill Book Company, Inc., New York, NY.
- Hahn, G. J., 1984, Experimental Design in the Complex World, *Technometrics*, **26**: 19–22.
- Hahn, G. J., and W. Q. Meeker, 1991, *Statistical Intervals: A Guide for Practitioners*, John Wiley & Sons, Inc., New York, NY.
- Hahn, G. J., and S. S. Shapiro, 1967, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York, NY.
- Hald, A., 1952a, *Statistical Tables and Formulas*, John Wiley & Sons, Inc., New York, NY.
- Hald, A., 1952b, *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York, NY.
- Hamada, M., A. Wilson, C. Reese, and H. Waller, 2008, *Bayesian Reliability*, Springer Publishing Company New York, NY.
- Hammersly, J. M., and D. C. Hanscomb, 1964, *Monte Carlo Methods*, Methuen, London, England.
- Harter, H. L., 1974–1976, The Method of Least Squares and Some Alternatives, *International Statistical Review*, Part I, **42**: 147–174; Part II, **42**: 235–264; Part III, **43**: 1–44; Part IV, **43**: 125–190; Part V, **43**: 269–278; Part VI, Subject and Author Indexes, **44**: 113–159.
- Hastie, T. J., and R. J. Tibshirani, 1990, *Generalized Additive Models*, Chapman and Hall, New York, NY.
- Hastings, N. A. J., and W. B. Peacock, 1975, *Statistical Distributions: A Handbook for Students and Practitioners*, A Halsted Press Book, John Wiley & Sons, Inc., New York, NY.
- Hicks, C. R., 1982, *Fundamental Concepts in the Design of Experiments*, 3rd Ed., Holt, Rinehart, and Winston, Austin, TX.
- Hines, W. W., D. C. Montgomery, D. M. Goldsman, and C. M. Borror, 1980, *Probability and Statistics in Engineering*, John Wiley & Sons, Inc., New York, NY.

- Hoaglin, D. C., F. Mosteller, and J. W. Tukey (Eds.), 1983, *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, Inc., New York, NY.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey (Eds.), 1985, *Exploring Data Tables, Trends, and Shapes*, John Wiley & Sons, Inc., New York, NY.
- Hoel, P. G., 1971, *Introduction to Mathematical Statistics*, 4th Ed., John Wiley & Sons, Inc., New York, NY.
- Hogan, H., 1992, The Post-Enumeration Survey: An Overview, *The American Statistician*, **46**, No. 4, pp. 261–269.
- Hollander, M., and D. A. Wolfe, 1973, *Nonparametric Statistical Methods*, John Wiley & Sons, Inc., New York, NY.
- Hooke, R., 1983, *How to Tell the Liars from the Statisticians*, Marcel Dekker, Inc., New York, NY.
- Hora, S. C., and W. J. Conover, 1981, *The Limiting Null Distribution of the F' Statistic in Factorial Designs with Rank-Score Transformed Data*, Texas Tech University Statistics Department Technical Report, Lubbock, TX.
- Howe, W. G., 1969, Two-sided Tolerance Limits for Normal Populations—Some Improvements, *Journal of the American Statistical Association*, **64**: 610–620.
- Huber, P. J., 1981, *Robust Statistics*, John Wiley & Sons, Inc., New York, NY.
- Huff, D., 1954, *How to Lie with Statistics*, W. W. Norton & Company, New York, NY.
- Hunter, J. S., 1975, The Technology of Quality, *RCA Engineer*, **30-3**: 8–15
- Iman, R. L., S. C. Hora, and W. J. Conover, 1981, *A Comparison of Asymptotically Distribution Free Procedures for the Analysis of Complete Blocks*, Texas Tech University Statistics Department Technical Report, Lubbock, TX.
- Isaaks, E. H., and R. M. Srivastava, 1990, *An Introduction to Applied Geostatistics*, Oxford University Press, Cary, NC.
- Jaech, J. L., 1973, *Statistical Methods in Nuclear Material Control*, TID-26298, National Technical Information Service, Springfield, VA.
- James, G., and R. C. James, 1951, *Mathematics Dictionary*, D. Van Nostrand Co., Inc., Princeton, NJ.
- Jaynes, E., 2003, *Probability Theory—The Logic of Science*, Cambridge University Press, Cambridge, England.
- Jeffreys, H., 1939, *Theory of Probability*, Oxford University Press, Oxford, England.
- Jeffreys, H., 1961, *Theory of Probability*, 3rd Ed., Clarendon Press, Gloucestershire, England.

- Johnston, J., 1984, *Econometric Methods*, 3rd Ed., McGraw-Hill Book Company, New York, NY.
- Johnson, M. E., 1987, *Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions*, John Wiley & Sons, Inc., New York, NY.
- Johnson, N. L., and S. Kotz, 1969, *Discrete Distributions*, Houghton Mifflin Co., Boston, MA.
- Johnson, N. L., and S. Kotz, 1970, *Continuous Univariate Distributions—I*, Houghton Mifflin Co., Boston, MA.
- Kasprzyk, D., G. Duncan, G. Kalton, and M. P. Singh (Eds.), 1989, *Panel Surveys*, John Wiley & Sons, Inc., New York, NY.
- Kelly, D. L., 1998, Bayesian Hypothesis Testing and the Maintenance Rule, *4th International Conference in Probabilistic Safety Assessment and Management (PSAM 4)*, Springer Publishing Company, New York, NY.
- Kemphorne, O., 1952, *The Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York, NY.
- Kendall, M. G., and W. B. Buckland, 1971, *A Dictionary of Statistical Terms*, 3rd Ed., Hafner Publishing Co., New York, NY.
- Kendall, M., A. Stuart, and J. K. Ord, 1987, *Kendall's Advanced Theory of Statistics*, 4th Ed., Volume 3: Design and Analysis, and Time-Series, Oxford University Press, New York, NY.
- Kendall, M., A. Stuart, and J. K. Ord, 1989, *Kendall's Advanced Theory of Statistics*, 5th Ed., Volume 1: Distribution Theory, Volume 2: Classical Inference and Relationship, Oxford University Press, New York, NY.
- Kennedy, W. J., Jr., and J. E. Gentle, 1980, *Statistical Computing*, Marcel Dekker, Inc., New York, NY.
- Klecka, W. R., 1980, *Discriminant Analysis*, Sage Publications, Newbury Park, CA.
- Kotz, S., and N. L. Johnson (Eds.), with C.B. Read, 1988, *Encyclopedia of Statistical Sciences* (in nine volumes), John Wiley & Sons, Inc., New York, NY.
- Krishnamoorthy, K., 2006, *Handbook of Statistical Distributions with Applications*, Chapman & Hall/CRC, Boca Raton, FL.
- Kruskal, W. H., and J. M. Tanur (Eds.), 1978, *International Encyclopedia of Statistics*, Volumes 1 and 2, The Free Press, New York, NY.

- Ku, H. H. (Ed.), 1969, *Precision Measurement and Calibration, Selected NBS Papers on Statistical Concepts and Procedures*, NBS Special Publication 300, Volume I, National Bureau of Standards, U.S. Government Printing Office, Washington, DC.
- LaBrecque, J., 1977, *Goodness-of-Fit Tests Based on Nonlinearity in Probability Plots*, *Technometrics*, **19**: 293–306.
- Lancaster, H. O., 1969, *The Chi-Squared Distribution*, John Wiley & Sons, Inc., New York, NY.
- Laplace, P. S., 1814, *A Philosophical Essay on Probabilities*, Dover Publications, Mineola, NY (reprint 1996).
- Large, D. W., and P. Michie, 1981, Proving that the Strength of the British Navy Depends on the Number of Old Maids in England: A Comparison of Scientific with Legal Proof, *Environmental Law*, **11**: 557–638.
- Lehmann, E. L., 1975, *Nonparametrics: Statistical Methods Based on Ranks*, Holden Day, San Francisco, CA.
- Levene, H. (1960). In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278–292.
- Levy, P. S., and S. Lemeshow, 1991, *Sampling of Populations: Methods and Applications*, John Wiley & Sons, Inc., New York, NY.
- Lieberman, G. J., and D. B. Owen, 1961, *Tables of the Hypergeometric Probability Distribution*, Stanford University Press, Stanford, CA.
- Lilienfeld, A. M., and D. E. Lilienfeld, 1980, *Foundations of Epidemiology*, Oxford University Press, Cary, NC.
- Lilliefors, H. W., 1967, On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *Journal of the American Statistical Association*, **62**: 399–402.
- Little, R. J. A., and D. B. Rubin, 1987, *Statistical Analysis with Missing Data*, John Wiley & Sons, Inc., New York, NY.
- Lwanga, S. K., and S. Lemeshow, 1991, *Sample Size Determination in Health Studies: A Practical Manual*, World Health Organization, Geneva, Switzerland.
- McDonald, P., 1971, The Analysis of a 2ⁿ Experiment by Means of Ranks, *J. Re Stat. Soc.*, Ser. C, 20: 259–275.
- Madansky, A., 1988, *Prescriptions for Working Statisticians*, Springer-Verlag, New York, NY.
- Mann, H.B., and D.R. Whitney, 1947, On a test whether one of two random variables is stochastically larger than another, *Annals of Mathematical Statistics*, 18, 50–60.

- Martz, H. F., and R. A. Waller, 1991, *Bayesian Reliability Analysis*, R. E. Krieger Publishing Co., Malabar, FL.
- Mason, R. L., R. F. Gunst, and J. L. Hess, 1989, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*, John Wiley & Sons, Inc., New York, NY.
- McCuen, R. H., 1985, *Statistical Methods for Engineers*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- McNemar, Q., 1947, Note on the Sampling Error of the Difference between Correlated Proportions or Percentages, *Psychometrika*, **12**: 153–157.
- Meyer, P. L., 1970, *Introductory Probability and Statistical Applications*, 2nd Ed., Addison-Wesley Publishing Company, Reading, MA.
- Meyer, S. L., 1975, *Data Analysis for Scientists and Engineers*, John Wiley & Sons, Inc., New York, NY.
- Miller, I., J. E. Freund, and R. A. Johnson, 1990, *Probability and Statistics for Engineers*, Prentice-Hall, Inc., Englewood Cliffs, NJ.
- Mood, A. M., and F. A. Graybill, 1963, *Introduction to the Theory of Statistics*, 2nd Ed., McGraw-Hill Book Company, Inc., New York, NY.
- Mood, A. M., F. A. Graybill, and D. C. Boes, 1974, *Introduction to the Theory of Statistics*, 3rd Ed., McGraw-Hill Book Company, Inc., New York, NY.
- Morrison, D. F., 2005, *Multivariate Statistical Methods*, Duxbury Press, Duxbury, MA.
- Mosteller, F., and J. W. Tukey, 1971, *Data Analysis and Regression*, Addison-Wesley Publishing Company, Reading, MA.
- Natrella, M. G., 1963, *Experimental Statistics*, National Bureau of Standards Handbook 91, U.S. Government Printing Office, Washington, DC.
- Nelson, C. R., 1973, *Applied Time Series Analysis for Managerial Forecasting*, Holden-Day, Inc., San Francisco, CA.
- Neter, J., W. Wasserman, C. J. Nachtsheim, and M. H. Kutner, 1996, *Applied Linear Statistical Models*, 4th Ed., Irwin, Homewood, IL.
- NIST/SEMATECH, *e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/>, 2006.
- NRC, 1989, *Nuclear Regulatory Commission Information Digest, 1992 Edition*, NUREG-1350, Volume 4, U.S. Nuclear Regulatory Commission, Washington, DC.
- NRC, 1994, *Applying Statistics*, NUREG 1475, rev 0, U.S. Government Printing Office, Washington, DC.

- NRC, 1998, *A nonparametric statistical methodology for the design and analysis of final status decommissioning surveys*, NUREG 1505, Rev. 1, U.S. Government Printing Office, Washington, DC.
- NRC, 2003, *Handbook of Parameter Estimation for Probabilistic Risk Assessment*, NUREG/CR-6823, U.S. Nuclear Regulatory Commission, Washington, DC.
- Odeh, R. E., and M. Fox, 1975, *Sample Size Choice: Charts for Experiments with Linear Models*, Marcel Dekker, Inc., New York, NY.
- Odeh, R. E., and D. B. Owen, 1980, *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, Marcel Dekker, Inc., New York, NY.
- Ordnance Corps, 1952, *Tables of the Cumulative Binomial Probabilities*, Ordnance Corps Pamphlet ORDP 20-1, U.S. Government Printing Office, Washington, DC.
- Ostle, B., and R. W. Mensing, 1979, *Statistics in Research*, 3rd Ed., Iowa State University Press, Ames, IA.
- Ott, L., and W. Mendenhall, 1984, *Understanding Statistics*, 4th Ed., Duxbury Press, Boston, MA.
- Owen, D. B., 1962, *Statistical Tables*, Addison-Wesley Publishing Co., Reading, MA.
- Pankratz, A., 1983, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, John Wiley & Sons, Inc., New York, NY.
- Pankratz, A., 1991, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, Inc., New York, NY.
- Parzen, E., 1960, *Modern Probability Theory and Its Applications*, John Wiley & Sons, Inc., New York, NY.
- Parzen, E., 1962, *Stochastic Processes*, Holden-Day, Inc., San Francisco, CA.
- Patel, J. K., C. H. Kapadia, and D. B. Owen, 1976, *Handbook of statistical distributions (Statistics, textbooks and monographs; v. 20)*, Marcel Dekker, Inc., New York, NY.
- Pearson, E. S., and H. O. Hartley, 1970 (Volume I) and 1976 (Volume II), *Biometrika Tables for Statisticians*, Cambridge University Press, Lowe and Brydone (Printers) Ltd., Thetford, Norfolk, England.
- Pearson, K., 1904, *On the Theory of Contingency and Its Relation to Association and Normal Correlation*, Drapers' Co. Memoires, Biometric Series No. 1, London, England.
- Plackett, R. L., 1981, *The Analysis of Categorical Data*, 2nd Ed., Macmillan Publishing Co., Inc., New York, NY.
- Press, S. J., 1988, *Bayesian Statistics: Principles, Models, and Applications*, John Wiley & Sons, Inc., New York, NY.

- Rand Corporation, 1955, *A Million Random Digits with 100,000 Normal Deviates*, Free Press of Glencoe, Glencoe, IL.
- Reichmann, W. J., 1971, *Use and Abuse of Statistics*, Oxford University Press, New York, NY.
- Romig, H. G., 1953, *50–100 Binomial Tables*, John Wiley & Sons, Inc., New York, NY.
- Rosner, B., 2010, *Fundamentals of Biostatistics*, 7th Ed., Brooks/Cole, Boston, MA.
- Rossi, P. H., J. D. Wright, and A. B. Anderson (Eds.), 1983, *Handbook of Survey Research*, Academic Press, Inc., Orlando, FL.
- Rubin, D. B., 1987, *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, Inc., New York, NY.
- Sandefur, J. T., 1993, *Discrete Dynamical Modeling*, Oxford University Press, Cary, NC.
- Satterthwaite, F. E., 1946, An Approximate Distribution of Estimates of Variance Components, *Biometrics Bulletin*, **2**: 110–114.
- Sawilowsky, S. S., G. C. Fahoome, (2003). *Statistics via Monte Carlo Simulation with Fortran*. Rochester Hills, MI: JMASM.
- Schilling, E. G., 1982, *Acceptance Sampling in Quality Control*, Marcel Dekker, Inc., New York, NY.
- Searle, S. R., 1987, *Linear Models for Unbalanced Data*, John Wiley & Sons, Inc., New York, NY.
- Shafer, G., 1990, The Unity and Diversity of Probability, *Statistical Science*, **5**, No. 4, pp. 435–462.
- Shapiro, S. S., and A. J. Gross, 1981, *Statistical Modeling Techniques*, Marcel Dekker, Inc., New York, NY.
- Shapiro, S. S., and M. B. Wilk, 1965, An Analysis of Variance Test for Normality (complete samples), *Biometrika*, **52**, 591–611.
- Shapiro, S. S., M. B. Wilk, and H. J. Chen, 1968, A Comparative Study of Various Tests for Normality, *Journal of the American Statistical Association*, **63**: 1343–1372.
- Sherr, T. S., 1972, *Attribute Sampling Inspection Procedure Based on the Hypergeometric Distribution*, WASH-1210, U.S. Atomic Energy Commission, U.S. Government Printing Office, Washington, DC.
- Shewhart, W. A., 1931, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Co., Inc., Princeton, NJ.
- Shooman, M. L., 1990, *Probabilistic Reliability: An Engineering Approach*, 2nd Ed., McGraw-Hill Book Company, Inc., New York, NY.

- Shreider, Y. A., 1966, *The Monte Carlo Method*, translated by G. J. Tee, Pergamon Press, New York, NY.
- Siegel, S., 1956, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, Inc., New York, NY.
- Siegel S. and N. J. Castellan, 1988, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, Inc., New York, NY.
- Slakter, M. J., 1965, A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Validity, *Journal of the American Statistical Society*, **60**: 854–858.
- Snedecor, G. W., and W. G. Cochran, 1989, *Statistical Methods*, 8th Ed., Iowa State University Press, Ames, IA.
- Staff of the Computational Laboratory, 1955, *Tables of the Cumulative Binomial Probability Distribution*, Harvard University Press, Cambridge, MA.
- Steel, R. G. D., and J. H. Torrie, 1980, *Principles and Procedures of Statistics: A Biometrical Approach*, 2nd Ed., McGraw-Hill Book Company, Inc., New York, NY.
- Stein, P., D. Coleman, and B. Gunter, 1986, Graphics for Display of Statistical Data, Supplement to *RCA Engineer*, **30**-3.
- Stiegler, S., 2008, Fisher and the 5% level, *Chance*, 21 (4).
- Stevens, S. S., 1946, On the Theory of Scales of Measurement, *Science*, **103**: 677–680.
- Stevens, W. L., 1939, Distribution of Groups in a Sequence of Alternatives, *Annals of Eugenics*, **9**: 10–17.
- Stiegler, S., 2008, Fisher and the 5% level, *Chance*, 21 (4).
- Swed, F. S., and C. Eisenhart, 1943, Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *Annals of Mathematical Statistics*, **14**: 66–87.
- Tabachnick, B. G., and L. S. Fidell, 1989, *Using Multivariate Statistics*, 2nd Ed., Harper & Row, Publishers, New York, NY.
- Tafte, E. R., 1977, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.
- Tanur, J., F. Mosteller, W. H. Kruskal, R. F. Link, R. S. Pieters, and G. R. Rising (Eds.), 1989, *Statistics: A Guide to the Unknown*, 3rd Ed., Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- Tietjen, G. L., and R. H. Moore, 1972, Some Grubbs-Type Statistics for the Detection of Several Outliers, *Technometrics*, **14**: 583–597.
- Toothaker, L. E., 1991, *Multiple Comparisons for Researchers*, Sage Publications, Inc., Newbury Park, CA.

- Tufte, E. R., 1983, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.
- Tufte, E. R., 1990, *Envisioning Information*, Graphics Press, Cheshire, CT.
- Tukey, J. W., 1977, *Exploratory Data Analysis*, Addison-Wesley Publishing Co., Reading, MA.
- Tukey, J. W., 1993, Graphic Comparisons of Several Aspects: Alternatives and Suggested Principles, *Journal of Computational and Graphical Statistics*, with comments by H. Wainer, discussion by R. A. Becker and W. S. Cleveland, and rejoinder by J. W. Tukey, **2**, 1, pp. 1–49.
- Upton, G. J. G., 1978, *The Analysis of Cross-Tabulated Data*, John Wiley & Sons, Inc., New York, NY.
- U.S. Code of Federal Regulations*, Title 10 (Energy), 2009.
- Velleman, P. F., and D. C. Hoaglin, 1981, *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press, Boston, MA.
- Velleman, P. F., and L. Wilkinson, 1993, Nominal, Ordinal, Interval, and Ratio Typologies are Misleading, *The American Statistician*, **47**, 1, pp. 65–72.
- Vesely, W. E., F. F. Goldberg, N. H. Roberts, and D. F. Haasl, 1981, *Fault Tree Handbook*, NUREG-0492, U.S. Nuclear Regulatory Commission, Washington, DC.
- Wadsworth, H. M. (Ed.), 1989, *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill Book Company, Inc., New York, NY.
- Wainer, H., 1984, How To Display Data Badly, *The American Statistician*, Vol. 38, No. 2, pp. 137–140.
- Wald, A., and J. Wolfowitz, 1940, On a Test Whether Two Samples Are from the Same Population, *Annals of Mathematical Statistics*, **11**: 147–162.
- Weinberg, G. H., and J. H. Shumaker, 1969, *Statistics, An Intuitive Approach*, 2nd Ed., Brooks/Cole Publishing Co., Monterey, CA.
- Weisberg, S., 1985, *Applied Linear Regression*, 2nd Ed., John Wiley & Sons, Inc., New York, NY.
- Welch, B. L., 1937, The Significance of the Difference between Two Means When the Population Variances Are Unequal, *Biometrika*, **29**: 350–362.
- Welch, B. L., 1947, The Generalization of “Student’s” Problem When Several Different Variances Are Involved, *Biometrika*, **34**: 28–35.
- Whicker, M. L., and L. Sigelman, 1991, *Computer Simulation Applications: An Introduction*, Sage Publications, Inc., Newbury Park, CA.
- Wilcoxon, F. and R. Wilcox, *Some Rapid Approximate Statistical procedure*, Lederle Laboratories, Pearl River, N.Y., 1964.

- Wilks, S. S., 1963, *Elementary Statistical Analysis*, Princeton University Press, Princeton, NJ.
- Williams, B., 1978, *A Sampler on Sampling*, John Wiley & Sons, Inc., New York, NY.
- Winer, B. J., 1971, *Statistical Principles in Experimental Design*, 2nd Ed., McGraw-Hill Book Company, Inc., New York, NY.
- Zwillinger, D. and S. Kokoska, 2000, *CRC Standard Probability and Statistics Tables and Formulae*, CRC Press, Boca Raton, FL.

Index

| | | | |
|---------------------------------------|----------|---|----------|
| acceptance number..... | 456 | census | 7 |
| acceptance sampling | 451 | central limit theorem | 124 |
| accuracy | 85 | central location | 17 |
| action limits | 446 | central tendency..... | 17 |
| addition rule for probabilities | 69 | central value | 17 |
| alarm limits | 446 | chain rule of probability | 76 |
| alternative hypothesis | 194 | Chebyshev's inequality | 37 |
| analysis of variance..... | 295, 315 | chi-square distribution..... | 129 |
| ANOVA..... | 295 | chi-square goodness-of-fit test .. | 206 |
| arithmetic mean | 17 | chi-square statistic | 206 |
| assurance-to-quality | 454 | class interval..... | 22 |
| attribute..... | 150, 404 | class marker..... | 22, 48 |
| average..... | 17 | classical statistical methods..... | 388 |
| axioms of probability | 67 | coefficient of determination | 382 |
| balanced design..... | 297 | column chart..... | 44 |
| balanced experiment | 269 | complement of event | 69 |
| bar chart..... | 44 | complementary cumulative
distribution function | 94 |
| Bartlett's test..... | 272 | complete factorial design..... | 316 |
| Bayes' factor | 399 | compound event | 64 |
| Bayes' Theorem..... | 78, 388 | conditional density | 102 |
| Bayesian inference | 79, 388 | conditional distribution..... | 100 |
| Behrens-Fisher problem..... | 288 | conditional probability | 73, 99 |
| Bernoulli distribution..... | 152 | confidence belt | 368 |
| Bernoulli trial..... | 152 | confidence bounds..... | 174 |
| bias..... | 86 | confidence coefficient | 174 |
| binary classification | 150 | confidence interval | 174 |
| binomial coefficient | 155 | confidence level..... | 174 |
| binomial distribution..... | 152 | confidence limits | 174 |
| binomial experiment | 413 | conjugate prior..... | 391 |
| block | 317 | consumer's risk | 197, 453 |
| blocked observations | 279 | continuous random variable | 91 |
| Box-Muller transformation..... | 520 | continuous scale | 6 |
| categorical scale..... | 6 | continuous variable | 5 |
| category | 230 | contrast | 106 |
| cell | 5 | | |

| | | | |
|---|---------|---|----------|
| control charts | 444 | distribution-free statistics | 462 |
| control limits | 444 | Dixon's test | 507 |
| correlation coefficient | 370 | domain of a variable | 104 |
| correlation, perfect | 370 | dot notation | 232 |
| covariance | 374 | dot plot | 49 |
| credible interval | 391 | Duncan's multiple range test | 309 |
| critical point | 195 | effect | 316 |
| critical region | 195 | elementary event | 64 |
| critical value | 195 | empirical cdf | 216 |
| cross-classification | 316 | empirical cumulative distribution
function | 216 |
| cumulative distribution
function (cdf) | 93 | empirical distribution function .. | 205 |
| curve fitting | 334 | empirical line | 337 |
| D'Agostino's D' test | 224 | empty set | 64 |
| data | 4 | equally likely events | 80 |
| data layout | 297 | error | 85 |
| data reduction | 16 | error factor (EF) | 127 |
| data transformation | 301 | error in regression model | 335 |
| data-based approach to
probability | 80 | error mean square | 349 |
| dataset | 4 | error sum of squares | 348 |
| datum | 4 | estimate | 172 |
| deciles | 26 | estimation | 170 |
| deductive reasoning | 3 | estimator | 171 |
| degrees of freedom | 29, 129 | event | 64 |
| density | 95 | expectation | 103 |
| dependent variable | 332 | expected count | 206, 233 |
| descriptive statistics | 16 | expected frequency | 207, 233 |
| deviation due to error | 345 | expected value, | 102 |
| deviation due to regression | 345 | experiment | 63 |
| dichotomous classification | 150 | experimental unit | 317 |
| discrete bivariate random
variable | 97 | exponential distribution | 138 |
| discrete distribution | 147 | exponential regression | 336 |
| discrete random variable | 90 | extrapolation | 366 |
| discrete uniform distribution | 148 | F ratio | 267, 306 |
| discrete variable | 5 | factor | 230, 316 |
| disjoint events | 65 | factorial design | 316 |
| distribution | 91 | failure frequency | 139 |
| | | failure rate | 140 |
| | | false alarm | 196 |

| | | | |
|-------------------------------------|----------|----------------------------------|----------|
| false negative | 197 | intercept..... | 333 |
| false positive | 196 | interquartile range | 26 |
| F-distribution | 134 | intersection | 64 |
| field..... | 5 | interval estimator..... | 174 |
| file..... | 5 | interval scale..... | 6 |
| Fisher's exact probability test ... | 239 | intra-class correlation | 375 |
| Fisher's Z-transformation | 378 | inventory difference (ID)..... | 250 |
| fixed effect..... | 318 | Jeffreys prior | 393 |
| fixed effects model | 297 | joint density | 101 |
| Fmax test..... | 269 | joint event..... | 70 |
| fractile..... | 26 | joint probability | 70 |
| Friedman's test | 489 | joint probability function..... | 97 |
| gamma distribution | 391 | Kendall's rank coefficient | 501 |
| Gaussian distribution | 113 | Kendall's tau | 501 |
| geometric distribution | 161 | Kolmogorov-Smirnov's test..... | 217 |
| geometric mean..... | 20 | Kruskal-Wallis' test | 482 |
| global correlation..... | 375 | K-S test..... | 217 |
| goodness-of-fit test | 203 | k-sample squared ranks test..... | 494 |
| grand total | 231 | law of total probability | 75 |
| Grubbs' tests | 508 | least squares..... | 342 |
| harmonic mean..... | 20, 309 | level | 230, 316 |
| Hartley's test..... | 269 | level of significance | 195 |
| hazard rate..... | 140 | Levene's test..... | 272 |
| heteroscedasticity..... | 269 | likelihood..... | 390 |
| histogram | 48 | limit | 446 |
| history (in simulation)..... | 514 | line plot..... | 45 |
| homoscedasticity..... | 269 | linear combination..... | 104 |
| Huff's criteria..... | 8 | linear contrast | 106 |
| hypergeometric distribution | 152, 404 | lognormal distribution | 126 |
| hypergeometric experiment | 154 | lot | 452 |
| hypothesis testing..... | 191 | lower confidence limit..... | 175 |
| imbalanced design | 297 | lower control limit (LCL)..... | 445 |
| inclusion-exclusion principle | 69 | main effect..... | 316 |
| independent random variable | 100 | Mann-Whitney's test..... | 475 |
| independent variable | 332 | marginal density | 101 |
| inductive reasoning..... | 3 | marginal probability | 70 |
| inference | 170 | marginal totals | 231 |
| informative prior | 389 | McNemar's test | 242 |
| interaction | 233, 316 | mean | 17, 103 |

| | | | |
|-----------------------------------|----------|-------------------------------------|--------|
| mean square error (MSE).. | 107, 350 | one-sided alternative hypotheses | 194 |
| mean time to failure | 140 | one-sided confidence intervals .. | 175 |
| mean time to repair | 140 | one-sided tolerance interval..... | 184 |
| measure of centrality..... | 17 | one-way ANOVA..... | 297 |
| measure of dispersion | 24 | one-way classification | 297 |
| measurement | 4, 84 | operating characteristic (OC) | 197 |
| measurement system..... | 85 | ordered pairs..... | 332 |
| median | 18 | ordinal scale..... | 6 |
| median test | 479 | outcome | 63 |
| metric scale | 6 | outlier | 504 |
| metrology..... | 84 | parameter..... | 7, 171 |
| midrange | 19 | parametric statistics | 462 |
| min cut upper bound | 77 | partition of sample space | 68 |
| minimal cut set..... | 77 | percentiles | 24 |
| minimum variance | 173 | pie chart..... | 41 |
| minimum variance unbiased | 173 | plus notation | 232 |
| mixed effects..... | 318 | point estimator..... | 172 |
| mode | 18 | Poisson distribution | 165 |
| modeling approach to | | Poisson experiment | 165 |
| probability..... | 80 | polynomial regression | 336 |
| Monte Carlo..... | 514 | pooling variances..... | 271 |
| multimodal..... | 18 | population..... | 7 |
| multiple comparison | 309 | population standard deviation | 28 |
| multiple linear regression.. | 336, 364 | posterior..... | 390 |
| multiple range test..... | 309 | posterior distribution | 390 |
| mutually exclusive | 65 | posterior probability | 79 |
| negative binomial distribution .. | 163 | power curve | 254 |
| negative binomial experiment... | 163 | power of the test | 252 |
| nominal scale | 6 | precision | 85 |
| noncritical region | 195 | prediction equation | 338 |
| non-informative prior..... | 389 | prediction interval | 367 |
| nonparametric statistics..... | 462 | prediction limits..... | 367 |
| normal density | 113 | prior | 389 |
| normal distribution..... | 113 | prior distribution..... | 389 |
| normalization constant..... | 390 | prior probability..... | 79 |
| null hypothesis | 194 | probabalistic risk assesment | |
| null set..... | 64 | (PRA)..... | 77 |
| observation..... | 4, 5 | probability | 62 |
| odds of an event..... | 399 | probability density function..... | 95 |

- probability function 67
 probability function, (pf)..... 148
 probability mass function (pmf) .91
 probability model..... 67
 probability of a Type I error 196
 probability of a Type II error 197
 probability of exceedance 196
 process control 443
 proposition 81
 pseudo standard deviation..... 310
 qualitative 4
 quality assurance..... 442
 quantiles..... 24
 quantitative 4
 quartiles 24
 quintiles 26
 random effect 297
 random effects 318
 random effects model 297
 random number 150
 random sample 105, 171
 random variable 90
 randomized complete block
 design..... 317
 range 24
 rank 216
 Rank ANOVA test 487
 rare event approximation 70
 ratio scale 6
 record 5
 rectangular distribution 110
 regression 334
 regression analysis 347
 regression line 337
 regression mean square 349
 regression sum of squares 347
 regression through the
 origin..... 337, 361
 rejection region 195
 relative frequency 80
 relative scale 6
 replication..... 317
 research hypothesis 194
 residual 345
 residual sum of squares 348
 response variable 316
 robust..... 269
 rounding rule 19
 row chart..... 44
 run 317, 447
 runs test 463
 sample 7, 171
 sample correlation coefficient ... 373
 sample mean 105
 sample points..... 64
 sample space..... 63
 sample variance 28
 sampling by attributes 443
 sampling by variables 443
 sampling with replacement..... 152
 sampling without replacement... 151
 Satterthwaite's approximation ... 288
 scale parameter 432
 scatter diagram 370
 scatter plot 370
 Shapiro-Wilk's test 221
 Shewhart charts 446
 signal, noise 302
 significant 192
 simple linear regression..... 336
 Simpson's paradox 241
 simulation 513
 skewness..... 33
 slope, intercept 333
 Spearman's rank correlation
 coefficient 498
 Spearman's rho..... 498
 spread 24

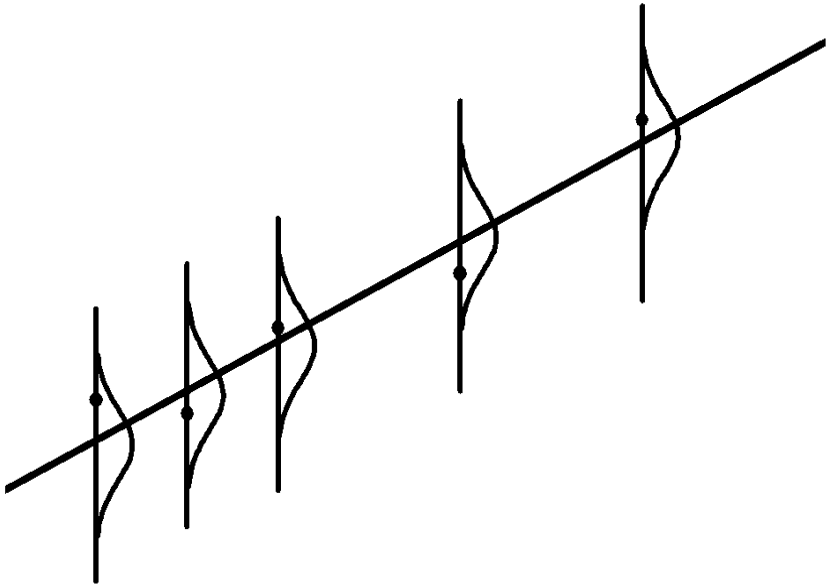
| | | | |
|-------------------------------------|-----|-------------------------------------|----------|
| squared ranks test..... | 491 | true value..... | 85 |
| standard deviation..... | 28 | truth table..... | 196 |
| standard error..... | 106 | two-sided confidence interval.... | 175 |
| standard error of mean differences | 107 | two-sided hypothesis..... | 194 |
| standard error of the mean..... | 106 | two-sided tolerance interval..... | 183 |
| standard normal distribution..... | 116 | Type I error..... | 196 |
| standard normal table..... | 118 | Type II error..... | 197 |
| standard uniform..... | 112 | unbiased..... | 86, 173 |
| standard uniform distribution.... | 112 | uncertainty..... | 87 |
| statistic..... | 7 | unconditional probability..... | 73 |
| statistical hypothesis..... | 194 | underlying normal..... | 126 |
| statistical model..... | 335 | uniform distribution..... | 110 |
| statistical significance..... | 192 | union..... | 64 |
| statistical tolerance limits..... | 183 | universal set..... | 64 |
| statistically dependent events..... | 76 | upper confidence bound..... | 175 |
| statistically independent events.. | 71 | upper confidence limit..... | 175 |
| step function..... | 216 | upper control limit (UCL)..... | 445 |
| Student's t-distribution..... | 132 | value..... | 4 |
| subjective probability approach.. | 81 | variability..... | 24 |
| success..... | 404 | variable..... | 4 |
| support..... | 104 | variance..... | 28 |
| systematic error..... | 86 | variance of a finite population.... | 28 |
| t-distribution..... | 132 | variance ratio..... | 267, 306 |
| test of hypothesis..... | 195 | variate..... | 516 |
| test statistic..... | 195 | Venn diagram..... | 65 |
| test-of-fit..... | 204 | warning limits..... | 446 |
| tolerance interval..... | 183 | weighted average..... | 271 |
| tolerance limit factors..... | 184 | weighted mean..... | 21 |
| tolerance limits..... | 183 | Welch approximation..... | 288 |
| total deviation..... | 345 | Wilcoxon's matched pairs test .. | 474 |
| total sum of squares..... | 347 | Wilcoxon's rank sum test..... | 475 |
| treatment group..... | 297 | within groups mean square..... | 306 |
| treatments..... | 297 | working hypothesis..... | 194 |
| trend..... | 349 | WRS test..... | 475 |
| trial..... | 317 | W-test..... | 221 |
| trimmed mean..... | 19 | zero-intercept model..... | 361 |
| trimodal..... | 18 | z-table..... | 118 |

| | | |
|---|--|--|
| NRC FORM 335
(9-2004)
NRCMD 3.7 | U.S. NUCLEAR REGULATORY COMMISSION | 1. REPORT NUMBER
(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.) |
| BIBLIOGRAPHIC DATA SHEET
(See instructions on the reverse) | | NUREG-1475, Rev. 1 |
| 2. TITLE AND SUBTITLE
Applying Statistics | 3. DATE REPORT PUBLISHED | |
| | MONTH
March | YEAR
2011 |
| | 4. FIN OR GRANT NUMBER | |
| 5. AUTHOR(S) | 6. TYPE OF REPORT | |
| D. Lurie, NRC (Ret.), L. Abramson, NRC (Ret.), J. Vail, NRC | Technical | |
| | 7. PERIOD COVERED (Inclusive Dates) | |
| 8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.) | | |
| Division of Risk Assessment
Office of Nuclear Reactor Regulation
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555 | | |
| 9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.) | | |
| Same as above. | | |
| 10. SUPPLEMENTARY NOTES | | |
| 11. ABSTRACT (200 words or less) | | |
| <p>Applying Statistics is intended to be both a reference book and a textbook on statistical methods. The material itself is organized into digestible segments, each exploring a particular method of applying statistics. These segments, in turn, are collected into chapters, each of which examines a set of linked statistical ideas. Many of the bite-sized segments contain examples of data which were encountered in or simulated from the application of statistics in the Nuclear Regulatory Commission. Because statistical ideas connect across all of our lives and interests, we also include a number of non-nuclear-regulation-related datasets.</p> <p>Chapters 1 through 8 introduce some fundamental statistical methods as an approach to dealing with everyday problems. Chapters 9 through 13 describe and expand upon "errors" as a statistical construct. Chapters 14 through 19 focus on multiple-group situations, addressing analysis-of-variance and simple linear regression and correlation methods. Chapters 20 through 23 deal with Bayesian statistics and basic ideas of probability and then apply them to hypergeometric, binomial, and Poisson processes. Chapters 24 through 27 discuss the important field of quality assurance and relatively knotty, easily misapplied, Nonparametric Statistics, Outliers, and Simulation applications.</p> | | |
| 12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.) | | 13. AVAILABILITY STATEMENT |
| Applying Statistics, Statistics, Graphical Displays, Probability, Errors, Random Variables, Continuous Distributions, Discrete Distributions, Statistical Estimation, Hypothesis Testing, Goodness of Fit Tests and Tests for Randomness, Contingency Tables, Testing Statistical Hypotheses: One Mean, Variances, Two Means, Several Means (One-way ANOVA), Beyond One-way Analysis of Variance, Regression, Correlation, Bayesian Statistics, Hypergeometric Experiments, Binomial Experiments, Poisson Experiments, Quality Assurance, Nonparametric Statistics, Outliers, Simulation. | | unlimited |
| | | 14. SECURITY CLASSIFICATION |
| | | <i>(This Page)</i>
unclassified |
| | | <i>(This Report)</i>
unclassified |
| | | 15. NUMBER OF PAGES |
| | | 16. PRICE |

(Continued from front flyleaf)

| Continuous distributions (continued) | | |
|---|---|--|
| Distribution | Density | Mean and variance |
| Exponential
(Section 7.13) | $f(y) = \lambda e^{-\lambda y}$
$y > 0$ | mean = $1/\lambda$
variance = $1/\lambda^2$ |
| Gamma
(Section 7.14) | $f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta}$,
$y > 0, \alpha > 0, \beta > 0$ | mean = α / β
variance = α / β^2 |
| Beta
(Section 7.15) | $f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}$,
$0 \leq y \leq 1, \alpha > 0, \beta > 0$ | mean = $\alpha / (\alpha + \beta)$
variance =
$\frac{\alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$ |

| Discrete distributions | | |
|---|---|---|
| Distribution | Probability function | Mean and variance |
| Discrete uniform
(Section 8.2) | $f(y; n) = 1/n,$ $y = 1, 2, \dots, n$ | mean = $(n+1)/2$
variance = $(n+1)(n-1)/12$ |
| Bernoulli
(Section 8.5) | $f(1) = Pr\{1\} = \pi$ $f(0) = Pr\{0\} = 1 - \pi$ | mean = π
variance = $\pi(1 - \pi)$ |
| Hyper-geometric
(Section 8.6, Chapter 21) | $f(y) = Pr\{y/N, M, n\}$ $= \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}},$ $y = \max(0, n + M - N), \dots, \min(M, n)$ | mean = nM/N
variance = $\left(\frac{nM}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$ |
| Binomial
(Section 8.7, Chapter 22)
Table T-8 | $f(y) = f(y; n, \pi) = Pr\{Y = y n, \pi\}$ $= \binom{n}{y} \pi^y (1 - \pi)^{n-y}$ $0 \leq \pi \leq 1, \quad y = 0, 1, \dots, n$ | mean = $n\pi$
variance = $n\pi(1 - \pi)$ |
| Geometric
(Section 8.8) | $f(y) = f(y; \pi) = Pr\{Y = y \pi\}$ $= \pi(1 - \pi)^{y-1}, \quad y = 1, 2, \dots$ | mean = $1/\pi$
variance = $(1 - \pi)/\pi^2$ |
| Negative binomial
(Section 8.9) | $f(y) = f(y s, \pi) = Pr\{y s, \pi\}$ $= \binom{y-1}{s-1} \pi^s (1 - \pi)^{y-s}$ $y = s, s+1, \dots$ | mean = s/π
variance = $s(1 - \pi)/\pi^2$ |
| Poisson
(Section 8.10, Chapter 23)
Table T-9 | $f(y) = f(y, \lambda) = Pr\{Y = y \lambda\}$ $= \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots$ | mean = λ
variance = λ |



Lurie
Abramson
Vail

Applying Statistics

Second
Edition



NUREG-1475
Rev. 1
March 2011