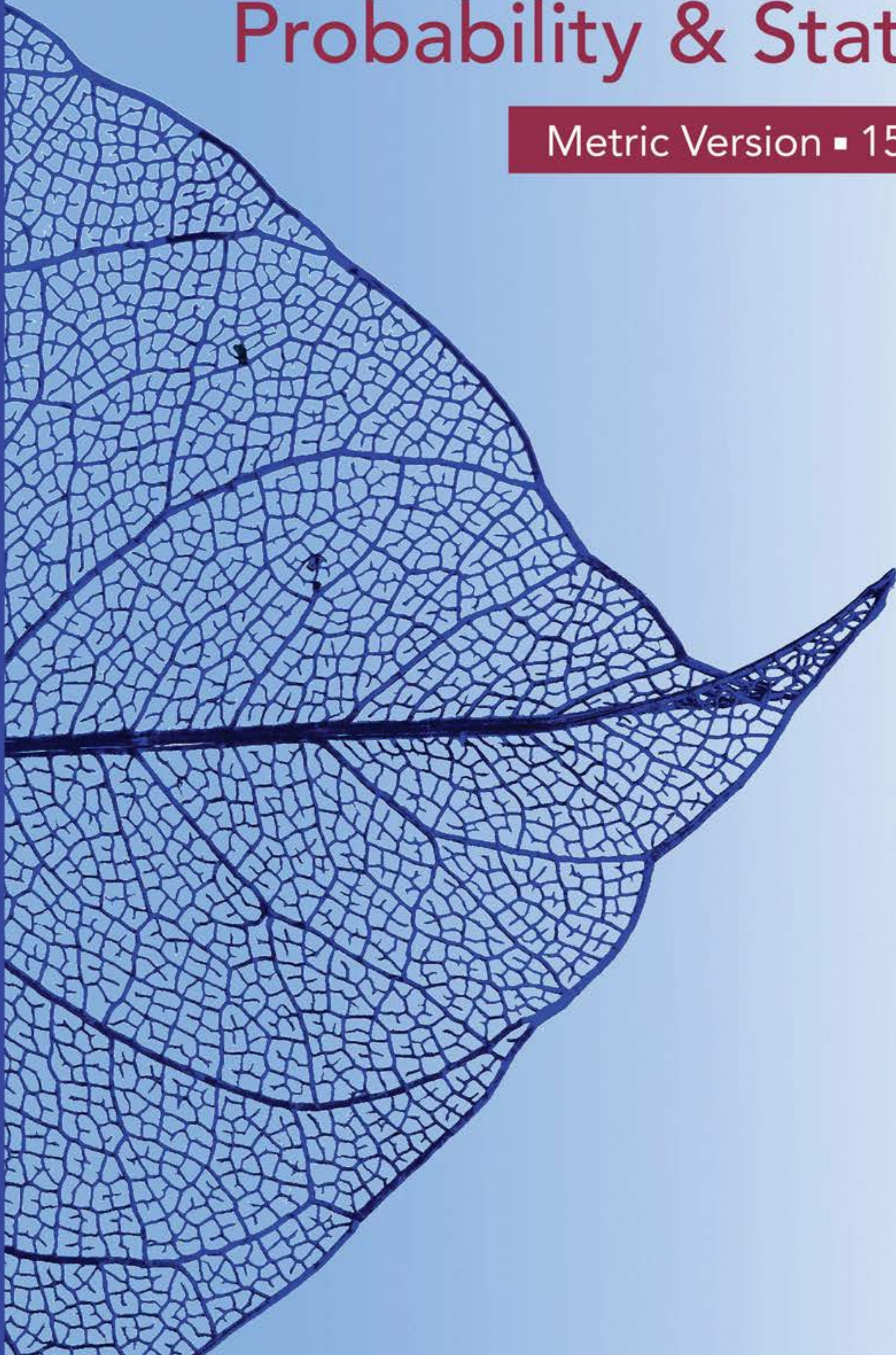
METRIC VERSION

Introduction to Probability & Statistics

Metric Version ▪ 15th Edition



Mendenhall ▪ Beaver ▪ Beaver

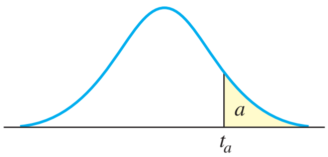


Table 4 Critical Values of *t*, pages 692–963

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	*df*
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
30	1.310	1.697	2.042	2.457	2.750	30
31	1.309	1.696	2.040	2.453	2.744	31
32	1.309	1.694	2.037	2.449	2.738	32
33	1.308	1.692	2.035	2.445	2.733	33
34	1.307	1.691	2.032	2.441	2.728	34
35	1.306	1.690	2.030	2.438	2.724	35
36	1.306	1.688	2.028	2.434	2.719	36
37	1.305	1.687	2.026	2.431	2.715	37
38	1.304	1.686	2.024	2.429	2.712	38
39	1.304	1.685	2.023	2.426	2.708	39
40	1.303	1.684	2.021	2.423	2.704	40
45	1.301	1.679	2.014	2.412	2.690	45
50	1.299	1.676	2.009	2.403	2.678	50
55	1.297	1.673	2.004	2.396	2.668	55
60	1.296	1.671	2.000	2.390	2.660	60
65	1.295	1.669	1.997	2.385	2.654	65
70	1.294	1.667	1.994	2.381	2.648	70
80	1.292	1.664	1.990	2.374	2.639	80
90	1.291	1.662	1.987	2.368	2.632	90
100	1.290	1.660	1.984	2.364	2.626	100
200	1.286	1.653	1.972	2.345	2.601	200
300	1.284	1.650	1.968	2.339	2.592	300
400	1.284	1.649	1.966	2.336	2.588	400
500	1.283	1.648	1.965	2.334	2.586	500
inf.	1.282	1.645	1.96	2.326	2.576	inf.

Source: Percentage points calculated using *Minitab* software.

Body temperature and heart rate 542
Breathing rates 74, 93, 441
Bulimia 392
Calcium 440, 460, 464
Calcium content 25
Cancer survivor rates 12, 187
Cerebral blood flow 226, 263, 391, 419
Chemical experiment 306, 515
Chemotherapy 642
Chicago weather 186
Chirping crickets 111, 523, 528
Chloroform 92
Cholesterol 393
Citrus red mite 325
Color preferences in mice 210
Cotton versus cucumber 580
Cure for insomnia 364
Cure for the common cold 358
Deep-sea research 617
Diabetes in children 208
Digitalis and calcium uptake 475
Disinfectants 402
Dissolved O2 content 392, 403, 459, 642
Drugs for hypertension 92
Drug potency 420
E. coli outbreaks 196
Early detection of breast cancer 363
Enzymes 401
Excedrin or Tylenol 318
FDA testing 175
Fossils 440
Fruit flies 136
Geothermal power 542
Genetic defects 233
Gestation times 121, 226, 523
Glucose tolerance 464
Good tasting medicine 619, 665
Ground or air 411
Gulf oil spill 48
Hazardous waste 26, 123
Healthy eating 358, 579
Healthy teeth 401, 411
Heart rate and exercise 465, 659
Hormone therapy and Alzheimer's disease 368
Human body temperatures 50, 95, 242, 264, 307, 313, 353, 359
Hungry rats 297
Impurities 428, 439
Iodine concentration 331
Jigsaw puzzles 654
Lead levels in blood 647
Lead levels in drinking water 358
Less red meat 321, 579
Lobsters 392, 541
Long stemmed roses 92
Lung cancer 233
Measurement error 263
Medical diagnostics 162
Mercury concentration in dolphins 84, 594
Metal corrosion and soil acids 572
Metabolism and weight gain 552
Monkey business 144
MRIs 164
Nematodes 549
Omega-3 Fats 250
Ore samples 74
Parasites in foxes 94
PCBs 377
Pearl millet 340
pH in rainfall 326
pH levels in water 660
Physical fitness 500, 612
Plant density 208
Plant genetics 157, 188, 234, 363
Plant science 537
Polluted seawater 67, 84
Pollution 326, 499, 678
Potassium levels 264
Potency of an antibiotic 353
Pulse rates 49, 227
Purifying organic compounds 392
Rain and snow 121
Recovery rates 647
Recurring illness 32, 91
Red blood cell count 25, 393
Rh factor 233, 285
Ring-necked pheasants 440
Runners and cyclists 402, 428, 443
San Andreas Fault 296
Screening tests 162
Sea urchins 440
Seed treatments 199
Selenium 311, 326
Shade or sun? 440
Slash pine seedlings 474
Sleep deprivation 515, 523
Smoking 331, 392
Sodium hydroxide 439
Spraying fruit trees 358
Sunflowers 227, 332
Survival times 32, 74, 85
Swampy sites 459, 464, 659
Sweet potato whitefly 363
Tai Chi and fibromyalgia 251, 368
Taste test for PTC 189
Tay-Sachs disease 188
Titanium 402
Toxic chemicals 664
Weights of babies 225, 263, 305, 352
Weights of turtles 642
Whitefly infestation 210, 499
White tailed deer 376

Social Sciences

Achievement scores 581
Achievement tests 75, 515, 550
Adolescents and social stress 376
Alcohol and altitude 442
American Presidents-age at death 26
Animation helps 502
Anxious infants 612
Back to work 17
Books or iPads? 401
Boomers, Xers and Millennial Men 375
Catching a cold 318, 321, 630
Choosing a mate 158
Discovery-based teaching 627
Drug offenders 156
Drug testing 156
Eye movement 642
Faculty salaries 263, 486, 501
Good at math? 460
Graduate teaching assistants 628
Hospital survey 143
Household size 101, 617
Images and word recall 251, 654
Intensive care 195
Jury duty 136
Laptops and learning 51, 524, 528
Math and art 677
Medical bills 189
Memory experiments 412
Midterm scores 118
Music in the workplace 412
No pass-no play rule for athletics 162
Organized religion 25
Political corruption 326
Preschool 33
Racial bias 250
Reaction times 410, 441, 442, 497, 498
Reducing hostility 458
Same-sex marriage 284, 306
SAT scores 92, 187, 313, 359, 376, 427
Smoking and cancer 157
Snacking and TV 242
Social ambivalence 92
Social Security numbers 74
Social skills training 110, 541, 671
Spending patterns 612
Starting salaries 312, 321, 359
Student ratings 671
Teaching biology 312
Test interviews 119, 515
Unbiased choices 144, 174, 199
Union Yes! 318
Violent crime 162
Want to be President? 16

EDITION
15

Introduction to Probability and Statistics

Metric Version

William Mendenhall, III
1925–2009

Robert J. Beaver
University of California, Riverside, Emeritus

Barbara M. Beaver
University of California, Riverside, Emerita

CENGAGE

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

CENGAGE

***Introduction to Probability and Statistics,* Fifteenth Edition, Metric Version**

William Mendenhall, III, Robert J. Beaver, Barbara M. Beaver

Metric Version prepared by Qaboos Imran

International Product Director: Timothy L. Anderson

Senior Product Assistant: Alexander Sham

Content Manager: Marianne Groth

Associate Marketing Manager: Tori Sitcawich

Associate Content Managers: Abby DeVeuve, Amanda Rose

Manufacturing Planner: Doug Bertke

IP Analyst: Reba Fredericks

IP Project Manager: Carly Belcher

Production Service/Compositor: SPi Global

Inventory Analyst: Sarah Ginsberg

Art Director: Vernon T. Boes

Senior Designer: Diana Graham

Cover Image: mikroman6/Getty Images

ISBN: 978-0-357-11446-9

Cengage International Offices

Asia
www.cengageasia.com
tel: (65) 6410 1200

Australia/New Zealand
www.cengage.com.au
tel: (61) 3 9685 4111

Brazil
www.cengage.com.br
tel: (55) 11 3665 9900

India
www.cengage.co.in
tel: (91) 11 4364 1111

Latin America
www.cengage.com.mx
tel: (52) 55 1500 6000

UK/Europe/Middle East/Africa
www.cengage.co.uk
tel: (44) 0 1264 332 424

Represented in Canada by Nelson Education, Ltd.
tel: (416) 752 9100 / (800) 668 0671
www.nelson.com

Cengage is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at: **www.cengage.com/global**.

For product information: **www.cengage.com/international**
Visit your local office: **www.cengage.com/global**
Visit our corporate website: **www.cengage.com**

Printed in the United States of America
Print Number: 01 Print Year: 2018

Brief Contents

INTRODUCTION: WHAT IS STATISTICS? 1

1 DESCRIBING DATA WITH GRAPHS 7

2 DESCRIBING DATA WITH NUMERICAL MEASURES 54

3 DESCRIBING BIVARIATE DATA 96

4 PROBABILITY 126

5 DISCRETE PROBABILITY DISTRIBUTIONS 167

6 THE NORMAL PROBABILITY DISTRIBUTION 212

7 SAMPLING DISTRIBUTIONS 245

8 LARGE-SAMPLE ESTIMATION 288

9 LARGE-SAMPLE TESTS OF HYPOTHESES 335

10 INFERENCE FROM SMALL SAMPLES 380

11 THE ANALYSIS OF VARIANCE 445

12 SIMPLE LINEAR REGRESSION AND CORRELATION 503

13 MULTIPLE LINEAR REGRESSION ANALYSIS 555

14 ANALYSIS OF CATEGORICAL DATA 599

15 NONPARAMETRIC STATISTICS 633

APPENDIX I 681

DATA SOURCES 714

ANSWERS TO SELECTED EXERCISES 727

INDEX 745

Contents

Introduction: What Is Statistics? 1

The Population and the Sample 3
Descriptive and Inferential Statistics 3
Achieving the Objective of Inferential Statistics: The Necessary Steps 4
Keys for Successful Learning 5

1 Describing Data with Graphs 7

1.1 Variables and Data 8
Types of Variables 9
Exercises 11
1.2 Graphs for Categorical Data 12
Exercises 15
1.3 Graphs for Quantitative Data 17
Pie Charts and Bar Charts 17
Line Charts 19
Dotplots 20
Stem and Leaf Plots 20
Interpreting Graphs with a Critical Eye 22
Exercises 24
1.4 Relative Frequency Histograms 27
Exercises 31

Chapter Review 35
Technology Today 35
Reviewing What You've Learned 47
CASE STUDY: How Is Your Blood Pressure? 53

2 Describing Data with Numerical Measures 54

Introduction 55
2.1 Measures of Center 55
Exercises 59
2.2 Measures of Variability 61
Exercises 66

2.3 Understanding and Interpreting the Standard Deviation 67
Tchebysheff's Theorem 67
The Empirical Rule 69
Approximating *s* Using the Range 71
Exercises 73
2.4 Measures of Relative Standing 76
z-Scores 76
Percentiles and Quartiles 77
The Five-Number Summary and the Box Plot 80
Exercises 83

Chapter Review 86
Technology Today 87
Reviewing What You've Learned 91
CASE STUDY: The Boys of Summer 95

3 Describing Bivariate Data 96

Introduction 97
3.1 Describing Bivariate Categorical Data 97
Exercises 99
3.2 Describing Bivariate Quantitative Data 101
Scatterplots 101
The Correlation Coefficient 104
The Least-Squares Line 106
Exercises 109

Chapter Review 112
Technology Today 112
Reviewing What You've Learned 118
CASE STUDY: Are Your Clothes *Really* Clean? 124

4 Probability 126

Introduction 127
4.1 Events and the Sample Space 127
Exercises 130
4.2 Calculating Probabilities Using Simple Events 131
Exercises 134
4.3 Useful Counting Rules 137
Using the *TI-83/84 Plus* Calculator 142
Exercises 142

4.4 Rules for Calculating Probabilities 144
Calculating Probabilities for Unions and Complements 146
Calculating Probabilities for Intersections 148
Exercises 155
4.5 Bayes' Rule 158
Exercises 161
Chapter Review 163
Reviewing What You've Learned 163
CASE STUDY: Probability and Decision Making in the Congo 166

5 Discrete Probability Distributions 167

5.1 Discrete Random Variables and Their Probability Distributions 168
Random Variables 168
Probability Distributions 168
The Mean and Standard Deviation for a Discrete Random Variable 170
Exercises 174
5.2 The Binomial Probability Distribution 176
Exercises 185
5.3 The Poisson Probability Distribution 189
Exercises 194
5.4 The Hypergeometric Probability Distribution 196
Exercises 198
Chapter Review 200
Technology Today 201
Reviewing What You've Learned 206
CASE STUDY: A Mystery: Cancers Near a Reactor 211

6 The Normal Probability Distribution 212

6.1 Probability Distributions for Continuous Random Variables 213
The Continuous Uniform Probability Distribution 215
The Exponential Probability Distribution 216
Exercises 217
6.2 The Normal Probability Distribution 218
The Standard Normal Random Variable 219
Calculating Probabilities for a General Normal Random Variable 222
Exercises 225
6.3 The Normal Approximation to the Binomial Probability Distribution 228
Exercises 232

Chapter Review 235
Technology Today 235
Reviewing What You've Learned 241
CASE STUDY: "Are You Going to Curve the Grades?" 244

7 Sampling Distributions 245

Introduction 246

7.1 Sampling Plans and Experimental Designs 246
Exercises 249

7.2 Statistics and Sampling Distributions 252
Exercises 254

7.3 The Central Limit Theorem and the Sample Mean 255
The Central Limit Theorem 255
The Sampling Distribution of the Sample Mean 258
Standard Error of the Sample Mean 259
Exercises 262

7.4 Assessing Normality 264

7.5 The Sampling Distribution of the Sample Proportion 268
Exercises 271

7.6 A Sampling Application: Statistical Process Control (Optional) 273
A Control Chart for the Process Mean: The $\overline{x}$ Chart 274
A Control Chart for the Proportion Defective: The p Chart 276
Exercises 278

Chapter Review 280
Technology Today 281
Reviewing What You've Learned 284
CASE STUDY: Sampling the Roulette at Monte Carlo 287

8 Large-Sample Estimation 288

8.1 Where We've Been and Where We're Going 289
Statistical Inference 289
Types of Estimators 290

8.2 Point Estimation 291
Exercises 296

8.3 Interval Estimation 298
Constructing a Confidence Interval 298
Large-Sample Confidence Interval for a Population Mean μ 300

Interpreting the Confidence Interval 301
Large-Sample Confidence Interval for a Population Proportion p 303
Using Technology 304
Exercises 304
8.4 Estimating the Difference Between Two Population Means 307
Exercises 311
8.5 Estimating the Difference Between Two Binomial Proportions 313
Using Technology 316
Exercises 316
8.6 One-Sided Confidence Bounds 319
Exercises 320
8.7 Choosing the Sample Size 322
Exercises 325
Chapter Review 326
Technology Today 327
Reviewing What You've Learned 330
CASE STUDY: How Reliable Is That Poll? *CBS News:* How and Where America Eats 333

9 Large-Sample Tests of Hypotheses 335

Introduction 336
9.1 A Statistical Test of Hypothesis 336
Exercises 339
9.2 A Large-Sample Test About a Population Mean 340
The Essentials of the Test 340
Calculating the p-Value 344
Two Types of Errors 348
The Power of a Statistical Test 349
Exercises 352
9.3 A Large-Sample Test of Hypothesis for the Difference Between Two Population Means 354
Hypothesis Testing and Confidence Intervals 356
Exercises 357
9.4 A Large-Sample Test of Hypothesis for a Binomial Proportion 360
Statistical Significance and Practical Importance 362
Exercises 363
9.5 A Large-Sample Test of Hypothesis for the Difference Between Two Binomial Proportions 365
Exercises 367

9.6 Concluding Comments on Testing Hypotheses 369

Chapter Review 370
Technology Today 371
Reviewing What You've Learned 375
CASE STUDY: An Aspirin a Day . . . ? 378

10 Inference from Small Samples 380

Introduction 381

10.1 Student's *t* Distribution 381
Assumptions behind Student's *t* Distribution 384
Exercises 385

10.2 Small-Sample Inferences Concerning a Population Mean 386
Exercises 390

10.3 Small-Sample Inferences for the Difference Between Two Population Means: Independent Random Samples 394
Exercises 400

10.4 Small-Sample Inferences for the Difference Between Two Means: A Paired-Difference Test 404
Exercises 409

10.5 Inferences Concerning a Population Variance 413
Exercises 419

10.6 Comparing Two Population Variances 421
Exercises 427

10.7 Revisiting the Small-Sample Assumptions 429

Chapter Review 430
Technology Today 431
Reviewing What You've Learned 439
CASE STUDY: School Accountability—Are We Doing Better? 443

11 The Analysis of Variance 445

11.1 The Design of an Experiment 446
Basic Definitions 446
What Is an Analysis of Variance? 447
The Assumptions for an Analysis of Variance 448
Exercises 448

11.2 The Completely Randomized Design: A One-Way Classification 449
Partitioning the Total Variation in the Experiment 450

Testing the Equality of the Treatment Means 453
Estimating Differences in the Treatment Means 455

Exercises 458

11.3 Ranking Population Means 461

Exercises 464

11.4 The Randomized Block Design: A Two-Way Classification 465
Partitioning the Total Variation in the Experiment 466
Testing the Equality of the Treatment and Block Means 469
Identifying Differences in the Treatment and Block Means 471
Some Cautionary Comments on Blocking 472

Exercises 473

11.5 The $a \times b$ Factorial Experiment: A Two-Way Classification 477
The Analysis of Variance for an $a \times b$ Factorial Experiment 479

Exercises 483

11.6 Revisiting the Analysis of Variance Assumptions 486
Residual Plots 487

11.7 A Brief Summary 489

Chapter Review 490
Technology Today 490
Reviewing What You've Learned 497
CASE STUDY: How to Save Money on Groceries! 502

12 Simple Linear Regression and Correlation 503

Introduction 504

12.1 Simple Linear Regression 504
A Simple Linear Model 505
The Method of Least Squares 507

Exercises 509

12.2 An Analysis of Variance for Linear Regression 511

Exercises 514

12.3 Testing the Usefulness of the Linear Regression Model 516
Inferences About β, the Slope of the Line of Means 516
The Analysis of Variance F-Test 519
Measuring the Strength of the Relationship: The Coefficient of Determination 520
Interpreting the Results of a Significant Regression 521

Exercises 522

12.4 Diagnostic Tools for Checking the Regression Assumptions 525
Dependent Error Terms 525
Residual Plots 525
Exercises 526
12.5 Estimation and Prediction Using the Fitted Line 530
Exercises 534
12.6 Correlation Analysis 537
Exercises 540

Chapter Review 543
Technology Today 544
Reviewing What You've Learned 549
CASE STUDY: Is Your Car "Made in the U.S.A."? 553

13 Multiple Linear Regression Analysis 555

Introduction 556
13.1 The Multiple Regression Model 556
13.2 Multiple Regression Analysis 558
The Method of Least Squares 558
The Analysis of Variance 559
Testing the Usefulness of the Regression Model 561
Interpreting the Results of a Significant Regression 562
Best Subsets Regression 563
Checking the Regression Assumptions 564
Using the Regression Model for Estimation and Prediction 564
Exercises 565
13.3 A Polynomial Regression Model 567
Exercises 570
13.4 Using Quantitative and Qualitative Predictor Variables in a Regression Model 573
Exercises 578
13.5 Testing Sets of Regression Coefficients 582
13.6 Other Topics in Multiple Linear Regression 584
Interpreting Residual Plots 584
Stepwise Regression Analysis 586
Binary Logistic Regression 587
Misinterpreting a Regression Analysis 587
13.7 Steps to Follow When Building a Multiple Regression Model 589

Chapter Review 589
Technology Today 590
Reviewing What You've Learned 592
CASE STUDY: "Made in the U.S.A."—Another Look 598

14 Analysis of Categorical Data 599
14.1 The Multinomial Experiment and the Chi-Square Statistic 600
14.2 Testing Specified Cell Probabilities: The Goodness-of-Fit Test 602
Exercises 604
14.3 Contingency Tables: A Two-Way Classification 606
The Chi-Square Test of Independence 607
Exercises 611
14.4 Comparing Several Multinomial Populations: A Two-Way Classification with Fixed Row or Column Totals 614
Exercises 616
14.5 Other Topics in Categorical Data Analysis 619
The Equivalence of Statistical Tests 619
Other Applications of the Chi-Square Test 620
Chapter Review 621
Technology Today 622
Reviewing What You've Learned 627
CASE STUDY: Who Is the Primary Breadwinner in Your Family? 631

15 Nonparametric Statistics 633
Introduction 634
15.1 The Wilcoxon Rank Sum Test: Independent Random Samples 634
Normal Approximation for the Wilcoxon Rank Sum Test 638
Exercises 641
15.2 The Sign Test for a Paired Experiment 643
Normal Approximation for the Sign Test 644
Exercises 646
15.3 A Comparison of Statistical Tests 648
15.4 The Wilcoxon Signed-Rank Test for a Paired Experiment 648
Normal Approximation for the Wilcoxon Signed-Rank Test 652
Exercises 653
15.5 The Kruskal–Wallis H-Test for Completely Randomized Designs 655
Exercises 658

15.6 The Friedman F_r-Test for Randomized Block Designs 660

Exercises 663

15.7 Rank Correlation Coefficient 666

Exercises 670

15.8 Summary 672

Chapter Review 672

Technology Today 673

Reviewing What You've Learned 676

CASE STUDY: Amazon HQ2 680

Appendix I 681

Table 1 Cumulative Binomial Probabilities 682

Table 2 Cumulative Poisson Probabilities 688

Table 3 Areas under the Normal Curve 690

Table 4 Critical Values of t 692

Table 5 Critical Values of Chi-Square 694

Table 6 Percentage Points of the F Distribution 696

Table 7 Critical Values of T for the Wilcoxon Rank Sum Test, $n_1 \leq n_2$ 704

Table 8 Critical Values of T for the Wilcoxon Signed-Rank Test, $n = 5(1)50$ 706

Table 9 Critical Values of Spearman's Rank Correlation Coefficient for a One-Tailed Test 707

Table 10 Random Numbers 708

Table 11 Percentage Points of the Studentized Range, $q_{.05}(k, df)$ 710

Data Sources 714

Answers to Selected Exercises 727

Index 745

Preface

Every time you pick up a newspaper or a magazine, watch TV, or scroll through Facebook, you encounter statistics. Every time you fill out a questionnaire, register at an online website, or pass your grocery rewards card through an electronic scanner, your personal information becomes part of a database containing your personal statistical information. You can't avoid it! In this digital age, data collection and analysis are part of our day-to-day activities. If you want to be an educated consumer and citizen, you need to understand how statistics are used and misused in our daily lives.

This international metric version is designed for classrooms and students outside of the United States. The units of measurement used in selected examples and exercises have been changed from U.S. Customary units to metric units. We did not update problems that are specific to U.S. Customary units, such as passing yards in football or data related to specific publications.

The Secret to Our Success

The first college course in introductory statistics that we ever took used *Introduction to Probability and Statistics* by William Mendenhall. Since that time, this text—currently in the fifteenth edition—has helped generations of students understand what statistics is all about and how it can be used as a tool in their particular area of application. The secret to the success of *Introduction to Probability and Statistics* is its ability to blend the old with the new. With each revision we try to build on the strong points of previous editions, and to look for new ways to motivate, encourage, and interest students using new technologies.

Hallmark Features of the Fifteenth Edition

The fifteenth edition keeps the traditional outline for the coverage of descriptive and inferential statistics used in previous editions. This revision maintains the straightforward presentation of the fourteenth edition. We have continued to simplify the language in order to make the text more readable—without sacrificing the statistical integrity of the presentation. We want students to understand how to apply statistical procedures, and also to understand

- how to meaningfully describe real sets of data
- how to explain the results of statistical tests in a practical way
- how to tell whether the assumptions behind statistical tests are valid
- what to do when these assumptions have been violated

Exercises

As with all previous editions, the variety and number of real applications in the exercise sets is a major strength of this edition. We have revised the exercise sets to provide new and interesting real-world situations and real data sets, many of which are drawn from current periodicals and journals. The fifteenth edition contains over 1900 exercises, many of which are new to this edition. Exercises are graduated in level of difficulty; some, involving only basic techniques, can be solved by almost all students, while others, involving practical applications and interpretation of results, will challenge students to use more sophisticated statistical reasoning and understanding. Exercises have been rearranged to provide a more even distribution of exercises within each chapter and a new numbering system has been introduced, so that numbering begins again with each new section.

Organization and Coverage

We believe that Chapters 1 through 10—with the possible exception of Chapter 3—should be covered in the order presented. The remaining chapters can be covered in any order. The analysis of variance chapter precedes the regression chapter, so that the instructor can present the analysis of variance as part of a regression analysis. Thus, the most effective presentation would order these three chapters as well.

Chapters 1–3 present descriptive data analysis for both one and two variables, using *MINITAB 18*, *Microsoft Excel 2016*®, and *TI-83/84 Plus* graphics. Chapter 4 includes a full presentation of probability. The last section of Chapter 4 in the fourteenth edition of the text, "Discrete Random Variables and Their Probability Distributions" has been moved to become the first section in Chapter 5. As in the fourteenth edition, the chapters on analysis of variance and linear regression include both calculational formulas and computer printouts in the basic text presentation. These chapters can be used with equal ease by instructors who wish to use the "hands-on" computational approach to linear regression and ANOVA and by those who choose to focus on the interpretation of computer-generated statistical printouts. This edition includes expanded coverage of the uniform and exponential distributions in Chapter 5 and normal probability plots for assessing normality in Chapter 7, in addition to an expanded t-table (Table 4 in Appendix I). New topics in Chapter 13 include best subsets regression procedures and binary logistic regression.

One important feature in the hypothesis testing chapters involves the emphasis on p-values and their use in judging statistical significance. With the advent of computer-generated p-values, these probabilities have become essential in reporting the results of a statistical analysis. As such, the observed value of the test statistic and its p-value are presented together at the outset of our discussion of statistical hypothesis testing as equivalent tools for decision-making. Statistical significance is defined in terms of preassigned values of α, and the *p-value approach* is presented as an alternative to the *critical value approach* for testing a statistical hypothesis. Examples are presented using both the *p-value* and *critical value* approaches to hypothesis testing. Discussion of the practical interpretation of statistical results, along with the difference between statistical significance and practical significance, is emphasized in the practical examples in the text.

Special Features of the Fifteenth Edition

- NEED TO KNOW. . .: This edition again includes highlighted sections called "NEED TO KNOW. . ." and identified by this icon. Need to Know... These sections provide information consisting of definitions, procedures, or step-by-step hints on problem solving for specific questions such as "NEED TO KNOW… How to Construct a Relative Frequency Histogram?" or "NEED TO KNOW… How to Decide Which Test to Use?"
- Graphical and numerical data description includes both traditional and EDA methods, using computer graphics generated by *MINITAB 18* for Windows and *MS Excel 2016*.
- Calculator screen captures from the *TI-84 Plus* calculator have been used for several examples, allowing students to access this option for data analysis.

Figure 9.9
TI-84 Plus output for Example 9.7

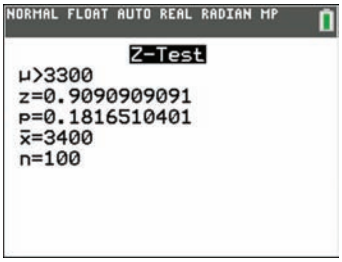
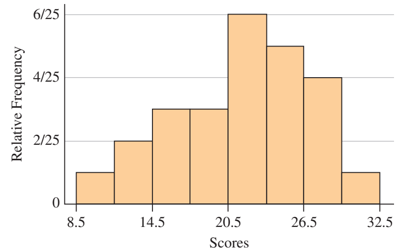


Figure 2.11
Relative frequency histogram for Example 2.8



E	F	G	H
Front Leg Room		*Rear Leg Room*	
Mean	41.9	Mean	28.350
Standard Error	0.221	Standard Error	0.409
Median	41.750	Median	28
Mode	41.500	Mode	28
Standard Deviation	0.699	Standard Deviation	1.292
Sample Variance	0.489	Sample Variance	1.669
Kurtosis	2.456	Kurtosis	-0.163
Skewness	1.353	Skewness	-0.021
Range	2.5	Range	4
Minimum	41	Minimum	26
Maximum	43.5	Maximum	30
Sum	419	Sum	283.5
Count	10	Count	10

- All examples and exercises in the text that contain printouts or calculator screen captures are based on *MINITAB 18*, *MS Excel 2016*, or the *TI-84 Plus* calculator. These outputs are provided for some exercises, while other exercises require the student to obtain solutions without using a computer.

Name	Length (km)	Name	Length (km)
Superior	560	Titicaca	195
Victoria	334	Nicaragua	163
Huron	330	Athabasca	333
Michigan	491	Reindeer	229
Aral Sea	416	Tonle Sap	112
Tanganyika	672	Turkana	246
Baykal	632	Issyk Kul	184
Great Bear	307	Torrens	208
Nyasa	576	Vänern	146
Great Slave	477	Nettilling	107
Erie	386	Winnipegosis	226
Winnipeg	426	Albert	160
Ontario	309	Nipigon	115
Balkhash	602	Gairdner	144
Ladoga	198	Urmia	144
Maracaibo	213	Manitoba	224
Onega	232	Chad	280
Eyre	144		

Source: The World Almanac and Book of Facts 2017

a. Use a stem and leaf plot to describe the lengths of the world's major lakes.

b. Use a histogram to display these same data. How does this compare to the stem and leaf plot in part a?

c. Are these data symmetric or skewed? If skewed, what is the direction of the skewing?

DATA SET DS0128 **6. Gulf Oil Spill Cleanup** On April 20, 2010, the United States experienced a major environmental disaster when a Deepwater Horizon drilling rig exploded in the Gulf of Mexico. The number of personnel and equipment used in the Gulf oil spill cleanup, beginning May 2, 2010 (Day 13) through June 9, 2010 (Day 51) is given in the following table.[17]

d. Use a bar graph to show the percentage of federal Gulf fishing areas closed.

e. Use a line chart to show the amounts of dispersants used. Is there any underlying straight line relationship over time?

DATA SET DS0129 **7. Election Results** The 2016 election was a race in which Donald Trump defeated Hillary Clinton and other candidates, winning 304 electoral votes, or 57% of the 538 available. However, Trump only won 46.1% of the popular vote, while Clinton won 48.2%. The popular vote (in thousands) for Donald Trump in each of the 50 states is listed as follows[18]:

AL	1319	HI	129	MA	1091	NM	320	SD	228
AK	163	ID	409	MI	2280	NY	2820	TN	1523
AZ	1252	IL	2146	MN	1323	NC	2363	TX	4685
AR	685	IN	1557	MS	701	ND	217	UT	515
CA	4484	IA	801	MO	1595	OH	2841	VT	95
CO	1202	KS	671	MT	279	OK	949	VA	1769
CT	673	KY	1203	NE	496	OR	782	WA	1222
DE	185	LA	1179	NV	512	PA	2971	WV	489
FL	4618	ME	336	NH	346	RI	181	WI	1405
GA	2089	MD	943	NJ	1602	SC	1155	WY	174

a. By just looking at the table, what shape do you think the distribution for the popular vote by state will have?

b. Draw a relative frequency histogram to describe the distribution of the popular vote for President Trump in the 50 states.

c. Did the histogram in part b confirm your guess in part a? Are there any outliers? How can you explain them?

TECHNOLOGY TODAY

The Role of Computers and Calculators in the Fifteenth Edition—Technology Today

Computers and scientific or graphing calculators are now common tools for college students in all disciplines. Most students are accomplished users of word processors, spreadsheets, and databases, and they have no trouble navigating through software packages in the Windows environment. Many own either a scientific or a graphing calculator, very often one of the many calculators made by Texas Instruments.™ We believe, however, that advances in computer technology should not turn statistical analyses into a "black box." Rather, we choose to use the computational shortcuts that modern technology provides to give us more time to emphasize statistical reasoning as well as the understanding and interpretation of statistical results.

In this edition, students will be able to use computers both for standard statistical analyses and as a tool for reinforcing and visualizing statistical concepts. Both *MS Excel 2016* and *MINITAB 18* are used exclusively as the computer packages for statistical analysis along with procedures available using the *TI-83* or *TI-84 Plus* calculators. However, we have chosen to isolate the instructions for generating computer and calculator output into individual sections called *Technology Today* at the end of each chapter. Each discussion uses numerical examples to guide the student through the *MS Excel* commands and option necessary for the procedures presented in that chapter, and then present the equivalent steps and commands needed to produce the same or similar results using *MINITAB* and the *TI-83/84 Plus*. We have included screen captures from *MS Excel*, *MINITAB 18*, and the *TI-84 Plus*, so that the student can actually work through these sections as "mini-labs."

If you do not need "hands-on" knowledge of *MINITAB*, *MS Excel*, or the *TI-83/84 Plus*, or if you are using another calculator or software package, you may choose to skip these sections and simply use the printouts as guides for the basic understanding of computer or calculator outputs.

TECHNOLOGY TODAY

Numerical Descriptive Measures in *Excel*

MS Excel provides most of the basic descriptive statistics presented in Chapter 2 using a single command on the **Data** tab. Other descriptive statistics can be calculated using the **Function Library** group on the **Formulas** tab.

EXAMPLE 2.17 The following data are the front and rear leg rooms (in inches) for 10 different compact sports utility vehicles[13]:

Make & Model	Front Leg Room	Rear Leg Room
Chevrolet Equinox	42.5	30.0
Ford Escape	41.5	28.0
Hyundai Tucson	41.5	28.0
Jeep Cherokee	43.5	30.0
Jeep Compass	41.5	28.0

Numerical Descriptive Measures in *MINITAB*

MINITAB provides most of the basic descriptive statistics presented in Chapter 2 using a single command in the drop-down menus.

EXAMPLE 2.18 The following data are the front and rear leg rooms (in inches) for 10 different compact sports utility vehicles[13]:

Make & Model	Front Leg Room	Rear Leg Room
Chevrolet Equinox	42.5	30.0
Ford Escape	41.5	28.0
Hyundai Tucson	41.5	28.0
Jeep Cherokee	43.5	30.0

Numerical Descriptive Measures on the *TI-83/84 Plus* Calculators

The *TI-83/84 Plus* calculators can be used to calculate descriptive statistics and create box plots using the **stat ➤ CALC** and **2nd ➤ stat plot** commands.

EXAMPLE 2.19 The following data are the front and rear leg rooms (in inches) for 10 different compact sports utility vehicles[13]:

Make & Model	Front Leg Room	Rear Leg Room
Chevrolet Equinox	42.5	30.0
Ford Escape	41.5	28.0

Study Aids

The many and varied exercises in the text provide the best learning tool for students embarking on a first course in statistics. The answers to all odd-numbered exercises are given in the back of the text. Each application exercise has a title, making it easier for students and instructors to immediately identify both the context of the problem and the area of application. All of the basic exercises have been rewritten and all of the applied exercises restructured according to increasing difficulty. New exercises have been introduced, dated exercises have been deleted, and a new numbering system has been introduced within each section.

The Basics

Normal Approximation? *Can the normal approximation be used to approximate probabilities for the binomial random variable x, with values for n and p given in Exercises 1–4? If not, is there another approximation that you could use?*

1. $n = 25$ and $p = .6$ **2.** $n = 45$ and $p = .05$

3. $n = 25$ and $p = .3$ **4.** $n = 15$ and $p = .5$

Using the Normal Approximation *Find the mean and standard deviation for the binomial random variable x*

12. $P(x \geq 6)$ and $P(x > 6)$ when $n = 15$ and $p = .5$

13. $P(4 \leq x \leq 6)$ when $n = 25$ and $p = .2$

14. $P(x \geq 7)$ and $P(x = 5)$ when $n = 20$ and $p = .3$

15. $P(x \geq 10)$ when $n = 20$ and $p = .4$

Applying the Basics

16. A *USA Today* snapshot found that 47% of Americans associate "recycling" with Earth Day.[9] Suppose a random sample of $n = 50$ adults are polled and that the

Students should be encouraged to use the "**NEED TO KNOW. . .**" sections as they occur in the text. The placement of these sections is intended to answer questions as they would normally arise in discussions. In addition, there are numerous hints called **"NEED A TIP?"** that appear in the margins of the text. The tips are short and concise.

● **Need a Tip?**
Parameter ⇔ Population
Statistic ⇔ Sample

In the previous three chapters, you have learned a lot about probability distributions, such as the binomial and normal distributions. The shape of the normal distribution is determined by its mean μ and its standard deviation σ, while the shape of the binomial distribution is determined by p. These numerical descriptive measures—called **parameters**—are needed to calculate the probability of observing sample results.

In practical situations, you may be able to decide which *type* of probability distribution to use as a model, but the values of the *parameters* that specify its *exact form* are unknown. Here are two examples:

Finally, sections called **Key Concepts and Formulas** appear in each chapter as a review in outline form of the material covered in that chapter.

CHAPTER REVIEW

Key Concepts and Formulas

I. Measures of the Center of a Data Distribution

1. Arithmetic mean (mean) or average
   a. Population: μ
   b. Sample of n measurements: $\bar{x} = \frac{\sum x_i}{n}$
2. Median; **position** of the median $= .5(n+1)$
3. Mode
4. The median may be preferred to the mean if the data are highly skewed.

2. The Empirical Rule can be used only for relatively mound-shaped data sets. Approximately 68%, 95%, and 99.7% of the measurements are within one, two, and three standard deviations of the mean, respectively.

IV. Measures of Relative Standing

1. Sample z-score: $z = \frac{x - \bar{x}}{s}$
2. pth percentile; p% of the measurements are smaller, and $(100 - p)$% are larger.
3. Lower quartile, Q_1; **position** of $Q_1 = .25(n+1)$

Instructor Resources

WebAssign

WebAssign for Mendenhall/Beaver/Beaver's *Introduction to Probability and Statistics,* 15th Edition, Metric Version is a flexible and fully customizable online instructional solution that puts powerful tools in the hands of instructors, empowering you to deploy assignments, instantly assess individual student and class performance, and help your students master the course concepts. With WebAssign's powerful digital platform and *Introduction to Probability and Statistics's* specific content, you can tailor your course with a wide range of assignment settings, add your own questions and content, and access student and course analytics and communication tools.

MindTap Reader

Available via WebAssign, MindTap Reader is Cengage's next-generation eBook. MindTap Reader provides robust opportunities for students to annotate, take notes, navigate, and interact with the text. Instructors can edit the text and assets in the Reader, as well as add videos or URLs.

Cognero

Cengage Learning Testing, powered by Cognero, is a flexible, online system that allows you to import, edit, and manipulate content from the text's Test Bank or elsewhere—including your own favorite test questions; create multiple test versions in an instant; and deliver tests from your LMS, your classroom, or wherever you want.

Instructor Solutions Manual

This time-saving online manual provides complete solutions to all the problems in the text. You can download the solutions manual from the Instructor Companion Website.

Instructor Companion Website

Everything you need for your course in one place! This collection of book-specific class tools is available online via www.cengage.com/login. Access and download PowerPoint presentations, images, Instructor Solutions Manual, data sets, and more.

SnapStat

Tell the story behind the numbers with SnapStat in WebAssign. Designed with students to bring stats to life, SnapStat uses interactive visuals to perform complex analysis online. Labs and Projects in WebAssign allow students to crunch their own data or choose from pre-existing data sets to get hands-on with technology and see for themselves that Statistics is much more than just numbers.

Student Resources

WebAssign

WebAssign for Mendenhall/Beaver/Beaver's *Introduction to Probability and Statistics*, 15th Edition, Metric Version lets you prepare for class with confidence. Its online learning platform for your math, statistics, and science courses helps you practice and absorb what you learn. Videos and tutorials walk you through concepts when you're stuck, and instant feedback and grading let you know where you stand—so you can focus your study time and perform better on in-class assignments. Study smarter with WebAssign!

MindTap Reader

Available via WebAssign, MindTap Reader is Cengage's next-generation eBook. MindTap Reader provides robust opportunities for students to annotate, take notes, navigate, and interact with the text. Annotations captured in MindTap are automatically tied to the Notepad app, where they can be viewed chronologically and in a cogent, linear fashion.

Online Technology Guides

Online Technology Guides, accessed via www.cengage.com, provide step-by-step instructions for completing problems using common statistical software.

SnapStat

Learn the story behind the numbers with SnapStat in WebAssign. Designed with students to bring stats to life, SnapStat uses interactive visuals to perform complex analysis online. Labs and Projects in WebAssign allow you to crunch your own data or choose from pre-existing data sets to get hands-on with technology and see for yourself that Statistics is much more than just numbers.

Acknowledgments

The authors are grateful to Catherine Van Der Laan and the editorial staff of Cengage Learning for their patience, assistance, and cooperation in the preparation of this edition.

Thanks are also due to fifteenth edition reviewers Olcay Akman, Matt Harris, Zhongming Huang, Bo Kai, Sarah Miller, and Katie Wheaton. We wish to thank authors and organizations for allowing us to reprint selected material; acknowledgments are made wherever such material appears in the text.

Robert J. Beaver
Barbara M. Beaver

Introduction
What Is Statistics?

Andrea Ricordi, Italy/Moment/Getty Images

What is statistics? Have you ever met a statistician? Do you know what a statistician does? Maybe you are thinking of the person who sits in the broadcast booth at the Rose Bowl, recording the number of pass completions, yards rushing, or interceptions thrown on New Year's Day. Or maybe just hearing the word *statistics* sends a shiver of fear through you. You might think you know nothing about statistics, but almost every time you turn on the news or scroll through your favorite news app, you will find statistics in one form or other! Here are some examples that we found just before the 2017 November elections:

- **Northam Heads Into Virginia Governor's Race With A Small Lead.** The first major statewide elections since President Trump was inaugurated take place on Tuesday…And while the race's final result by itself isn't likely to tell us much about the national political environment, it *is likely* to have a big effect on the 2018 midterms. Polls show a fairly close race, with Northam slightly favored to win [over Ed Gillespie]. An average of the last 10 surveys give Northam a 46 percent-to-43 percent advantage. Over the past month, there has been a tightening of the race, with Gillespie closing what had been a 6-point lead. In the individual polls, though, there is a fairly wide spread. Northam has led by as much as 17 percentage points (a Quinnipiac University survey) and has trailed by as much as 8 points (a Hampton University poll).[1]

—www.fivethirtyeight.com

- **Why Trump Has a Lock on the 2020 GOP Nomination.** In interviews with nearly three-dozen GOP strategists and fundraisers over the past several tumultuous weeks, virtually everyone told me that…they expect Trump to coast to the GOP nomination in 2020…the hurdles to a 2020 primary challenge are vivid when considering a recent *Washington Post/ABC News* poll that found 91% of Trump voters said they'd vote for him again…This *ABC News/Washington Post* poll was conducted by landline and cellular telephone Oct. 29-Nov. 1, 2017, in English and Spanish, among a random national sample of 1005 adults. Results have a margin of sampling error of 3.5 points, including the design effect.[2]

—www.cnn.com

Articles similar to these can be found in all forms of news media, and, just before a presidential or congressional election, a new poll is reported almost every day. These articles are very familiar to us; however, they might leave you with some unanswered questions. How were the people in the poll selected? Will these people give the same response tomorrow? Will they give the same response on election day? Will they even vote? Are these people representative of all those who will vote on election day? It is the job of a statistician to ask these questions and to find answers for them in the language of the poll.

Most Believe "Cover-Up" of JFK Assassination Facts

A majority of the public believes the assassination of President John F. Kennedy was part of a larger conspiracy, not the act of one individual. In addition, most Americans think there was a cover-up of facts about the 1963 shooting. Almost 50 years after JFK's assassination, a FOX news poll shows many Americans disagree with the government's conclusions about the killing. The **Warren Commission** found that **Lee Harvey Oswald** acted alone when he shot Kennedy, but 66 percent of the public today think the assassination was "part of a larger conspiracy" while only 25 percent think it was the "act of one individual."

"For older Americans, the Kennedy assassination was a traumatic experience that began a loss of confidence in government," commented Opinion Dynamics President John Gorman. "Younger people have grown up with movies and documentaries that have pretty much pushed the 'conspiracy' line. Therefore, it isn't surprising there is a fairly solid national consensus that we still don't know the truth."

(The poll asked): "Do you think that we know all the facts about the assassination of President John F. Kennedy or do you think there was a cover-up?"

	We Know All the Facts (%)	There Was a Cover-Up	(Not Sure)
All	14	74	12
Democrats	11	81	8
Republicans	18	69	13
Independents	12	71	17

—www.foxnews.com[3]

When you see an article like this one, do you simply read the title and the first paragraph, or do you read further and try to understand the meaning of the numbers? How did the authors get these numbers? Did they really interview every American with each political affiliation? It is the job of the statistician to answer some of these questions.

Hot News: 98.6°F Not Normal

After believing for more than a century that 98.6°F was the normal body temperature for humans, researchers now say normal is not normal anymore.

For some people at some hours of the day, 99.9°F could be fine. And readings as low as 96°F turn out to be highly human.

The 98.6°F standard was derived by a German doctor in 1868. Some physicians have always been suspicious of the good doctor's research. His claim: 1 million readings—in an epoch without computers.

So Mackowiak & Co. took temperature readings from 148 healthy people over a three-day period and found that the mean temperature was 98.2°F. Only 8 percent of the readings were 98.6°F.

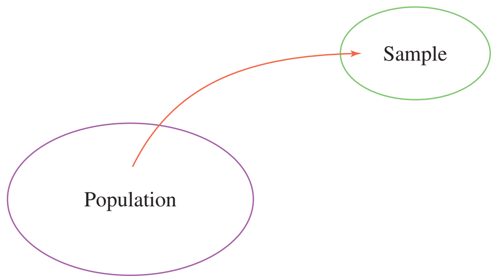
—*The Press-Enterprise*[4]

What questions do you have when you read this article? How did the researcher select the 148 people, and how can we be sure that the results based on these 148 people are accurate when applied to the general population? How did the researcher arrive at the normal "high" and "low" temperatures given in the article? How did the German doctor record 1 million temperatures in 1868? This is another statistical problem with an application to everyday life.

Statistics is a branch of mathematics that has applications in almost every part of our daily life. It is a new and unfamiliar language for most people, however, and, like any new language, statistics can seem overwhelming at first glance. But once the language of statistics is learned and understood, it provides a powerful tool for data analysis in many different fields of application.

The Population and the Sample

In the language of statistics, one of the most basic concepts is **sampling.** In most statistical problems, a specified number of measurements or data—a **sample**—is drawn from a much larger body of measurements, called the **population.**



For the body-temperature experiment, the sample is the set of body-temperature measurements for the 148 healthy people chosen by the experimenter. We hope that the sample is representative of a much larger body of measurements—the population—the body temperatures of all healthy people in the world!

Which is more important to us, the sample or the population? In most cases, we are interested primarily in the population, but identifying each member of the population may be difficult or impossible. Imagine trying to record the body temperature of every healthy person on earth or the presidential preference of every registered voter in the United States! Instead, **we try to describe or predict the behavior of the population on the basis of information obtained from a representative sample from that population.**

The words *sample* and *population* have two meanings for most people. For example, you read that a Gallup poll conducted in the United States was based on a sample of 1823 people. Presumably, each person interviewed is asked a particular question, and that person's response represents a single measurement in the sample. Is the sample the set of 1823 people, or is it the 1823 responses that they give?

In statistics, we distinguish between the set of objects on which the measurements are taken and the measurements themselves. To experimenters, the objects on which measurements are taken are called **experimental units.** The sample survey statistician calls them **elements of the sample.**

Descriptive and Inferential Statistics

When first presented with a set of measurements—whether a sample or a population—you need to find a way to organize and summarize it. The branch of statistics that gives us tools for describing sets of measurements is called **descriptive statistics.** You have seen descriptive statistics in many forms: bar charts, pie charts, and line charts presented by a political candidate; numerical tables in the media; or the average rainfall amounts on your favorite weather app. Computer-generated graphics and numerical summaries are commonplace in our everyday communication.

DEFINITION

Descriptive statistics are procedures used to summarize and describe the important characteristics of a set of measurements.

If the set of measurements is the entire population, you need only to draw conclusions based on the descriptive statistics. However, it might be too expensive or too time consuming to identify each member of the population. Maybe listing the entire population would destroy it—for example, measuring the amount of force required to cause a football helmet crack. For these or other reasons, you may have only a sample from the population. By looking at the sample, you want to answer questions about the population as a whole. The branch of statistics that deals with this problem is called **inferential statistics.**

DEFINITION

Inferential statistics are procedures used to make inferences (that is, draw conclusions, make predictions, make decisions) about a population from information contained in a sample drawn from this population.

The **objective of inferential statistics** is to make inferences about a population from information contained in a sample.

Achieving the Objective of Inferential Statistics: The Necessary Steps

How can you make inferences about a population using information contained in a sample? The task becomes simpler if you organize the problem into a series of logical steps.

1. **Specify the questions to be answered and identify the population of interest.** In the Virginia election poll, the objective is to determine who will get the most votes on election day. So, the population of interest is the set of all votes in the Virginia election. When you select a sample, it is important that the sample be representative of *this* population, not the population of voter preferences on some day prior to the election.
2. **Decide how to select the sample.** This is called the *design of the experiment* or the *sampling procedure*. Is the sample representative of the population of interest? For example, if a sample of registered voters is selected from the city of San Francisco, will this sample be representative of all voters in California? Will it be the same as a sample of "likely voters"—those who are likely to actually vote in the election? Is the sample large enough to answer the questions posed in step 1 without wasting time and money on additional information? A good sampling design will answer the questions posed with minimal cost to the experimenter.
3. **Select the sample and analyze the sample information.** No matter how much information the sample contains, you must use an appropriate method of analysis to obtain it. Many of these methods, which depend on the sampling procedure in step 2, are explained in the text.

4. **Use the information from step 3 to make an inference about the population.** Many different procedures can be used to make this inference, and some are better than others. For example, 10 different methods might be available to estimate human response to an experimental drug, but one procedure might be more accurate than others. You should use the best inference-making procedure available (many of these are explained in the text).
5. **Determine the reliability of the inference.** Since you are using only a fraction of the population in drawing the conclusions described in step 4, you might be wrong! If an agency conducts a statistical survey for you and estimates that your company's product will gain 34% of the market this year, how much confidence can you place in this estimate? Is this estimate accurate to within 1, 5, or 20 percentage points? Is it reliable enough to be used in setting production goals? Every statistical inference should include a measure of reliability that tells you how much confidence you have in the inference.

Now that you have learned a few basic terms and concepts, we again pose the question asked at the beginning of this discussion: Do you know what a statistician does? The statistician's job is to carry out all of the preceding steps.

Keys for Successful Learning

As you begin to study statistics, you will find that there are many new terms and concepts to be mastered. Since statistics is an applied branch of mathematics, many of these basic concepts are mathematical—developed and based on results from calculus or higher mathematics. However, you do not have to be able to prove the results in order to apply them in a logical way. In this text, we use numerical examples and commonsense arguments to explain statistical concepts, rather than more complicated mathematical arguments.

Computers and calculators are now readily available to many students and provide them with an invaluable tool. In the study of statistics, even the beginning student can use packaged programs to perform statistical analyses with a high degree of speed and accuracy. Some of the more common statistical packages available at computer facilities are *MINITAB*™, SAS, and SPSS. Personal computers and laptops will support *MINITAB, MS EXCEL*, *JMP,* and others. Many students are familiar with the *TI-83* or *TI-84 Plus* calculators, that have many built-in statistics functions. There are even online statistical programs and interactive "applets" that students can use.

These programs, called **statistical software,** differ in the types of analyses available, the options within the programs, and the forms of printed results (called **output**). However, they are all similar. In this book, we use both *MINITAB* and *Microsoft Excel* as statistical tools. Understanding the basic output of these packages will help you interpret the output from other software systems. Similarly, understanding the results shown on your *TI-83* or *TI-84 Plus* calculator will make understanding a different calculator much easier.

At the end of most chapters, you will find a section called "*Technology Today.*" These sections present numerical examples to guide you through the *MINITAB*, *MS Excel*, and *TI-83/84 Plus* commands and options that are used for the procedures in that chapter. If you are using *MINITAB*, *MS Excel*, or your *TI-83/84 Plus* calculator in a lab or home setting, you may want to work through this section using your own computer or calculator so that you become familiar with the hands-on methods. If you do not need hands-on knowledge of *MINITAB*, *MS Excel*, or the *TI-83/84 Plus*, you may choose to skip this section and simply use the computer printouts or calculator screen captures for analysis as they appear in the text.

Most important, using statistics successfully requires common sense and logical thinking. For example, if we want to find the average height of all students at a particular university, would we select our entire sample from the members of the basketball team? In the body-temperature example, the logical thinker would question an 1868 average based on 1 million measurements—when computers had not yet been invented.

As you learn new statistical terms, concepts, and techniques, remember to view every problem with a critical eye and be sure that the rule of common sense applies. Throughout the text, we will remind you of the pitfalls and dangers in the use or misuse of statistics. Benjamin Disraeli once said that there are three kinds of lies: *lies, damn lies,* and *statistics*! Our purpose is to prove this claim to be wrong—to show you how to make statistics *work* for you and not *lie* for you!

As you continue through the book, refer back to this introduction every once in a while. Each chapter will increase your knowledge of statistics and should, in some way, help you achieve one of the steps described here. Each of these steps is important in achieving the overall objective of inferential statistics: to make inferences about a population using information contained in a sample drawn from that population.

Describing Data with Graphs

1

© Photographee.eu/Shutterstock.com

How Is Your Blood Pressure?

Is your blood pressure normal, or is it too high or too low? The case study at the end of this chapter examines a large set of blood pressure data. You will use graphs to describe these data and compare your blood pressure with that of others of your same age and gender.

LEARNING OBJECTIVES

Many sets of measurements are samples selected from larger populations. Other sets constitute the entire population, as in a national census. In this chapter, you will learn what a *variable* is, how to classify variables into several types, and how measurements or data are generated. You will then learn how to use graphs to describe data sets.

CHAPTER INDEX

- Data distributions and their shapes (1.1, 1.3)
- Dotplots (1.3)
- Pie charts, bar charts, line charts (1.2, 1.3)
- Qualitative and quantitative variables—discrete and continuous (1.1)
- Relative frequency histograms (1.4)
- Stem and leaf plots (1.3)
- Univariate and bivariate data (1.1)
- Variables, experimental units, samples and populations, data (1.1)

● Need to Know…

How to Construct a Stem and Leaf Plot

How to Construct a Relative Frequency Histogram

1.1 Variables and Data

In Chapters 1 and 2, we will present some basic techniques in *descriptive statistics*—the branch of statistics concerned with describing sets of measurements, both *samples* and *populations*. Once you have collected a set of measurements, how can you display this set in a clear, understandable, and readable form? First, you must be able to define what is meant by measurements or "data" and to categorize the types of data that you are likely to encounter in real life. We begin by introducing some definitions.

DEFINITION

A **variable** is a characteristic that changes or varies over time and/or for different individuals or objects under consideration.

For example, body temperature is a variable that changes over time within a single individual; it also varies from person to person. Religious affiliation, ethnic origin, income, height, age, and number of offspring are all variables—characteristics that vary depending on the individual chosen.

In the Introduction, we defined an *experimental unit* or an *element of the sample* as the object on which a measurement is taken. This is the same as saying that an *experimental unit* is the object on which a *variable* is measured. When a variable is actually measured on a set of experimental units, a set of measurements or **data** result.

DEFINITION

An **experimental unit** is the individual or object on which a variable is measured. A single **measurement** or data value results when a variable is actually measured on an experimental unit.

If a measurement is obtained for every experimental unit in the entire collection, the resulting data set constitutes the *population* of interest. Any smaller subset of measurements is a *sample*.

DEFINITION

A **population** is the set of all measurements of interest to the investigator.

DEFINITION

A **sample** is a subset of measurements selected from the population of interest.

EXAMPLE 1.1

A set of five students is selected from all undergraduates at a large university, and measurements are entered into a spreadsheet as shown in Figure 1.1. Identify the various elements involved in obtaining this set of measurements.

Solution The *experimental unit* on which the variables are measured is a particular undergraduate student on the campus, found in column A. Five *variables* are measured for each student: grade point average (GPA), gender, year in college, major, and current number of units

Figure 1.1
Measurements on five undergraduate students

	A	B	C	D	E	F
1	Student	GPA	Gender	Year	Major	Number of Units
2	1	2	F	Fr	Psychology	16
3	2	2.3	F	So	Mathematics	15
4	3	2.9	M	So	English	17
5	4	2.7	M	Fr	English	15
6	5	2.6	F	Jr	Business	14

Sheet 1

enrolled. Each of these characteristics varies from student to student. If we consider the GPAs of all students at this university to be the population of interest, the five GPAs in column B represent a *sample* from this population. If the GPA of each undergraduate student at the university had been measured, we would have the entire *population* of measurements for this variable.

The second variable measured on the students is gender, in column C. This variable is somewhat different from GPA, because it typically takes one of two values—male (M) or female (F). If we could identify each member of the population, it would consist of a set of Ms and Fs, one for each student at the university. The third and fourth variables, year and major, also involve nonnumerical data—year has four categories (Fr, So, Jr, Sr), and major has one category for each undergraduate major on campus. The last variable, current number of units enrolled, is numerically valued, consisting of a set of numbers rather than a set of qualities or characteristics.

Although we have discussed each variable individually, remember that we have measured each of these five variables on a single experimental unit: the student. Therefore, in this example, a "measurement" really consists of five observations, one for each of the five measured variables. For example, the measurement taken on student 2 produces this observation:

(2.3, F, So, Mathematics, 15)

There is a difference between a *single* variable measured on a single experimental unit and *multiple* variables measured on a single experimental unit as in Example 1.1.

DEFINITION

Univariate data results when a single variable is measured on a single experimental unit.

DEFINITION

Bivariate data results when two variables are measured on a single experimental unit. **Multivariate data** results when more than two variables are measured.

If you measure the body temperatures of 148 people, the resulting data are *univariate.* In Example 1.1, five variables were measured on each student, resulting in *multivariate* data.

Types of Variables

Variables can be classified into one of two types: **qualitative** or **quantitative.**

DEFINITION

Qualitative variables measure a quality or characteristic on each experimental unit. **Quantitative variables** measure a numerical quantity or amount on each experimental unit.