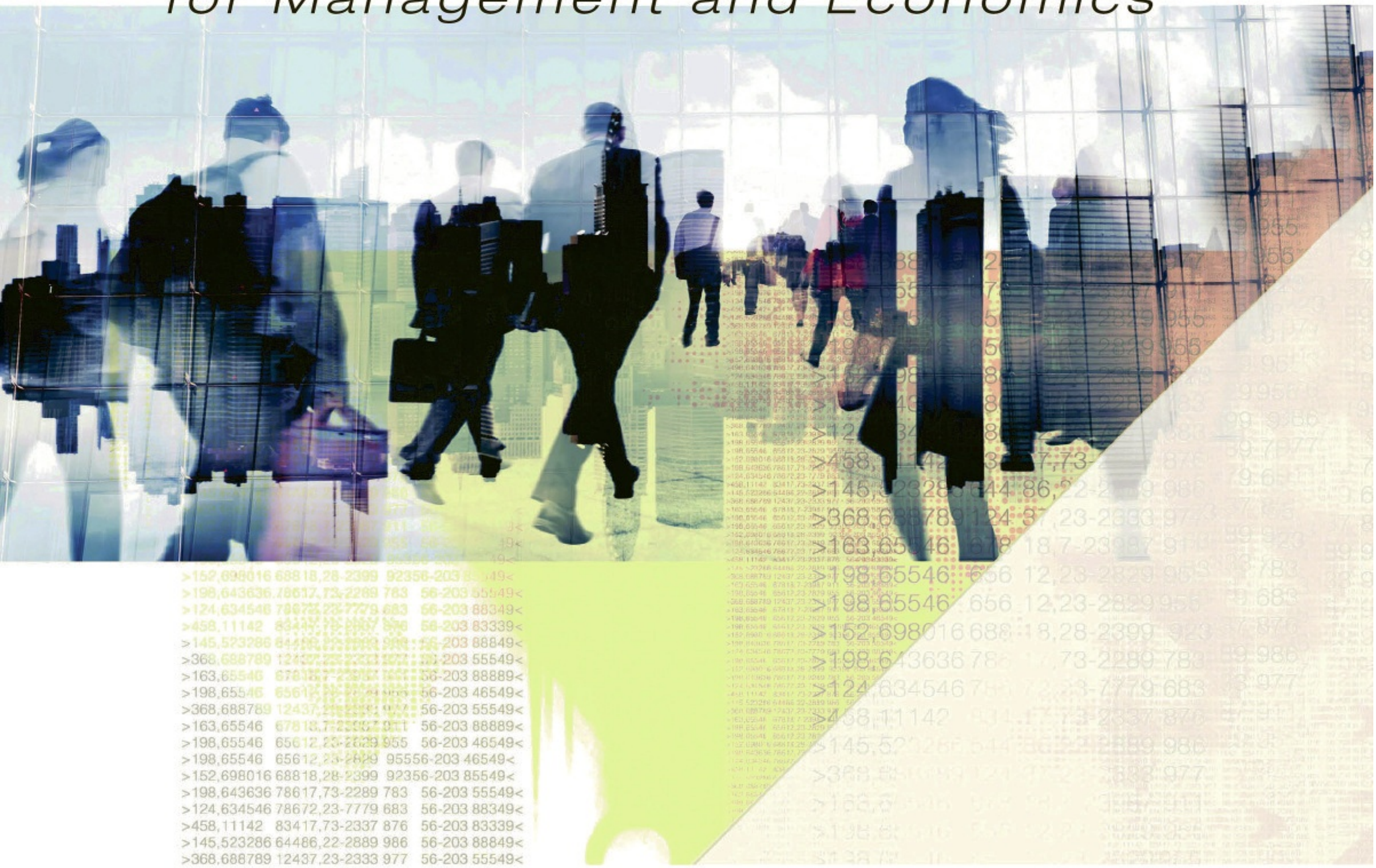
GERALD
KELLER

STATISTICS

for Management and Economics



IDENTIFY COMPUTE INTERPRET

11e

A GUIDE TO STATISTICAL TECHNIQUES

Problem Objectives

DATA TYPES	Describe a Population	Compare Two Populations	Compare Two or More Populations	Analyze Relationship between Two Variables	Analyze Relationship among Two or More Variables
Interval	Histogram **Section 3-1** Line chart **Section 3-2** Mean, median, and mode **Section 4-1** Range, variance, and standard deviation **Section 4-2** Percentiles and quartiles **Section 4-3** t-test and estimator of a mean **Section 12-1** Chi-squared test and estimator of a variance **Section 12-2**	Equal-variances t-test and estimator of the difference between two means: independent samples **Section 13-1** Unequal-variances t-test and estimator of the difference between two means: independent samples **Section 13-1** t-test and estimator of mean difference **Section 13-3** F-test and estimator of ratio of two variances **Section 13-4** Wilcoxon rank sum test **Section 19-1** Wilcoxon signed rank sum test **Section 19-2**	One-way analysis of variance **Section 14-1** LSD multiple comparison method **Section 14-2** Tukey's multiple comparison method **Section 14-2** Two-way analysis of variance **Section 14-4** Two-factor analysis of variance **Section 14-5** Kruskal-Wallis test **Section 19-3** Friedman test **Section 19-3**	Scatter diagram **Section 3-3** Covariance **Section 4-4** Coefficient of correlation **Section 4-4** Coefficient of determination **Section 4-4** Least squares line **Section 4-4** Simple linear regression and correlation **Chapter 16** Spearman rank correlation **Section 19-4**	Multiple regression **Chapters 17 & 18**
Nominal	Frequency distribution **Section 2-2** Bar chart **Section 2-2** Pie chart **Section 2-2** z-test and estimator of a proportion **Section 12-3** Chi-squared goodness-of-fit test **Section 15-1**	z-test and estimator of the difference between two proportions **Section 13-5** Chi-squared test of a contingency table **Section 15-2**	Chi-squared test of a contingency table **Section 15-2**	Chi-squared test of a contingency table **Section 15-2**	Not covered
Ordinal	Median **Section 4-1** Percentiles and quartiles **Section 4-3**	Wilcoxon rank sum test **Section 19-1** Sign test **Section 19-2**	Kruskal-Wallis test **Section 19-3** Friedman test **Section 19-3**	Spearman rank correlation **Section 19-4**	Not covered

GENERAL SOCIAL SURVEY AND SURVEY OF CONSUMER FINANCES EXERCISES

Chapter	GSS	Page	SCF	Page
2	2.66–2.73	43	2.74–2.78	44
3	3.23–3.27	59	3.28–3.31	59
	3.72–3.78	73	3.79–3.81	73
4	4.19–4.22	95	4.23–4.26	96
	4.49–4.52	104	4.53–4.56	105
	4.77–4.79	109	4.80–4.82	110
	4.145–4.150	135	4.151–4.154	135
12	12.54–12.58	387	12.59–12.69	388
	12.117–12.123	411	12.124–12.129	411
	12.171–12.178	421	12.179–12.190	422
13	13.54–13.68	453	13.69–13.80	454
	13.106–13.109	472	13.110–13.111	472
	13.121–13.122	479	13.123–13.126	479
	13.167–13.184	497	13.185–13.193	498
	13.235–13.257	506	13.258–13.260	508
	A13.18–A13.28	515	A13.29–A13.38	515
14	14.23–14.46	534	14.47–14.56	536
	14.67–14.80	545	14.81–14.82	546
	14.98–14.99	557		
	A14.19–A14.31	589	A14.32–A14.42	590
15	15.21–15.23	599	15.24–15.25	600
	15.53–15.62	611	15.63–15.78	612
	A15.17–A15.32	627	A15.33–A15.43	629
16	16.50–16.65	663	16.66–16.75	664
	16.101–16.112	670		
	16.142–16.150	681		
	A16.17–A16.31	689	A16.32–A16.41	690
17	17.18–17.21	711	17.22–17.24	712
	17.43–17.46	715		
	A17.17–A17.35	730	A17.36–A17.43	731
18	18.24–18.33	750		
	18.40–18.42	758		
19	19.17–19.30	777	19.31–19.34	778
	19.58–19.59	793		
	19.85–19.93	804		
	19.107–19.117	812		
	A19.26–A19.42	829	A19.42–A19.49	830

APPLICATION SECTIONS

APPLICATION SUBSECTION

Statistics
for Management and Economics

11e

GERALD KELLER

CENGAGE Learning

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

CENGAGE Learning

Statistics for Management and Economics, Eleventh Edition
Gerald Keller

Vice President, General Manager, Social Science & Qualitative Business: Erin Joyner

Sr. Product Team Manager: Joe Sabatino

Sr. Product Manager: Aaron Arnsparger

Content Developer: Conor Allen

Product Assistant: Renee Schnee

Sr. Marketing Director: Kristen Hurd

Sr. Marketing Manager: Nate Anderson

Sr. Content Project Manager: Martha Conway

Manufacturing Planner: Ron Montgomery

Production Service: SPi Global

Sr. Art Director: Michelle Kunkler

Cover and Internal Designer: cmillerdesign

Cover Image: © Rawpixel.com/Shutterstock.com

Intellectual Property Analyst: Brittani Morgan

Intellectual Property Project Manager: Reba Frederics

Library of Congress Control Number: 2017932174

Package:
ISBN: 978-1-337-09345-3

Book Only:
ISBN: 978-1-337-29694-6

Loose-leaf Edition:
ISBN: 978-1-337-29876-6

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Printed in the United States of America
Print Number: 01 Print Year: 2017

BRIEF CONTENTS

CONTENTS

PREFACE

Businesses are increasingly using statistical techniques to convert data into information. For students preparing for the business world, it is not enough merely to focus on mastering a diverse set of statistical techniques and calculations. A course and its attendant textbook must provide a complete picture of statistical concepts and their applications to the real world. ***Statistics for Management and Economics*** is designed to demonstrate that statistical methods are vital tools for today's managers and economists.

Fulfilling this objective requires the several features that I have built into this book. First, I have included data-driven examples, exercises, and cases that demonstrate statistical applications that are and can be used by marketing managers, financial analysts, accountants, economists, operations managers, and others. Many are accompanied by large and genuine data sets. Second, I reinforce the applied nature of the discipline by teaching students how to choose the correct statistical technique. Third, I teach students the concepts that are essential to interpret the statistical results.

Why I Wrote This Book

Business is complex and requires effective management to succeed. Managing complexity requires many skills. There are more competitors, more places to sell products, and more places to locate workers. As a consequence, effective decision-making is more crucial than ever before. On the other hand, managers have more access to larger and more detailed data that are potential sources of information. However, to achieve this potential requires that managers know how to convert data into information. This knowledge extends well beyond the arithmetic of calculating statistics. Unfortunately, this is what most textbooks offer—a series of unconnected techniques illustrated mostly with manual calculations. This continues a pattern that goes back many years. What is required now is a complete approach to applying statistical techniques.

When I started teaching statistics in 1971, books demonstrated how to calculate statistics and, in some cases, how various formulas were derived. One reason for doing so was the belief that by doing calculations by hand, students would be able to understand the techniques and concepts. When the first edition of this book was published in 1988, an important goal was to teach students to identify the correct technique. Through the next 10 editions, I refined my approach to emphasize interpretation and decision-making equally. I now divide the solution of statistical problems into three stages and include them in every appropriate example: (1) *identify* the technique, (2) *compute* the statistics, and (3) *interpret* the results. The compute stage can be completed in any or all of three ways: manually (with the aid of a calculator), using Excel, and using XLSTAT. For those courses that wish to use the computer extensively, manual calculations can be played down or omitted completely. Conversely, those that wish to emphasize manual calculations may easily do so, and the computer solutions can be selectively introduced or skipped entirely. This approach is designed to provide maximum flexibility, and it leaves to the instructor the decision of if and when to introduce the computer.

I believe that my approach offers several advantages:

- An emphasis on identification and interpretation provides students with practical skills that they can apply to real problems they will face regardless of whether a course uses manual or computer calculations.
- Students learn that statistics is a method of converting data into information. With 1136 data files and corresponding problems that ask students to interpret statistical results, students are given ample opportunities to practice data analysis and decision-making.
- The optional use of the computer allows for larger and more realistic exercises and examples.

Placing calculations in the context of a larger problem allows instructors to focus on more important aspects of the decision problem. For example, more attention needs to be devoted to interpret statistical results. Proper interpretation of statistical results requires an understanding of the probability and statistical concepts that underlie the techniques and an understanding of the context of the problems. An essential aspect of my approach is teaching students the concepts. I do so by providing Excel worksheets that allow students to perform "what-if" analyses. Students can easily see the effect of changing the components of a statistical technique, such as the effect of increasing the sample size.

Efforts to teach statistics as a valuable and necessary tool in business and economics are made more difficult by the positioning of the statistics course in most curricula. The required statistics course in most undergraduate programs appears in the first or second year. In many graduate programs, the statistics course is offered in the first semester of a three-semester program and the first year of a two-year program. Accounting, economics, finance, human resource management, marketing, and operations management are usually taught after the statistics course. Consequently, most students will not be able to understand the general context of the statistical application. This deficiency is addressed in this book by "Applications in …" sections, subsections, and boxes. Illustrations of statistical applications in businesses that students are unfamiliar with are preceded by an explanation of the background material.

- For example, to illustrate graphical techniques, we use an example that compares the histograms of the returns on two different investments. To explain what financial analysts look for in the histograms requires an understanding that risk is measured by the amount of variation in the returns. The example is preceded by an "Applications in Finance" box that discusses how return on investment is computed and used.
- Later when I present the normal distribution, I feature another "Applications in Finance" box to show why the standard deviation of the returns measures the risk of that investment.
- Forty-two application boxes are scattered throughout the book.

Some applications are so large that I devote an entire section or subsection to the topic. For example, in the chapter that introduces the confidence interval estimator of a proportion, I also present market segmentation. In that section, I show how the confidence interval estimate of a population proportion can yield estimates of the sizes of market segments. In other chapters, I illustrate various statistical techniques by showing how marketing managers can apply these techniques to determine the differences that exist between market segments. There are five such sections and one subsection in this book.

The "Applications in …" segments provide great motivation to the student who asks, "How will I ever use this technique?"

New in This Edition

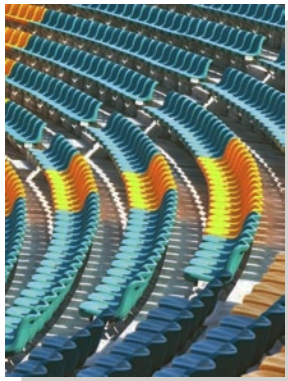
The data from the last eight General Social Surveys and the last four Surveys of Consumer Finances have been included, which produced 528 new exercises. Students will have the opportunity to convert real data into information. Instructors can use these data sets to create hundreds of additional examples and exercises.

Many of the examples, exercises, and cases using real data in the 10th edition have been updated. These include the data on wins, payrolls, and attendance in baseball, basketball, football, and hockey; returns on stocks listed on the New York Stock Exchange, NASDAQ, and Toronto Stock Exchange; and global warming.

I've created many new examples and exercises. Here are the numbers for the 11th edition: 142 solved examples, 2460 exercises, 32 cases, 1136 data sets, 31 appendixes containing 37 solved examples, 98 exercises, and 25 data sets, for a grand total of 179 worked examples, 2558 exercises, 32 cases, and 1161 data sets.

Data Driven: The Big Picture

Solving statistical problems begins with a problem and data. The ability to select the right method by problem objective and data type is **a valuable tool for business**. Because business decisions are driven by data, students will leave this course equipped with the tools they need to make effective, informed decisions in all areas of the business world.



lzf/Shutterstock.com

Identify the Correct Technique

Examples introduce the first crucial step in this three-step (*identify–compute–interpret*) approach. Every example's solution begins by examining the data type and problem objective and then identifying the right technique to solve the problem.

EXAMPLE 13.1*

DATA Xm13-01

Direct and Broker-Purchased Mutual Funds

Millions of investors buy mutual funds (see page 178 for a description of mutual funds), choosing from thousands of possibilities. Some funds can be purchased directly from banks or other financial institutions whereas others must be purchased through brokers, who charge a fee for this service. This raises the question, Can investors do better by buying mutual funds directly than by purchasing mutual funds through brokers? To help answer this question, a group of researchers randomly sampled the annual returns from mutual funds that can be acquired directly and mutual funds that are bought through brokers and recorded the net annual returns, which are the returns on investment after deducting all relevant fees. These are listed next.

Direct					Broker				
9.33	4.68	4.23	14.69	10.29	3.24	3.71	16.4	4.36	9.43
6.94	3.09	10.28	−2.97	4.39	−6.76	13.15	6.39	−11.07	8.31
16.17	7.26	7.1	10.37	−2.06	12.8	11.05	−1.9	9.24	−3.99
16.97	2.05	−3.09	−0.63	7.66	11.1	−3.12	9.49	−2.67	−4.44
5.94	13.07	5.6	−0.15	10.83	2.73	8.94	6.7	8.97	8.63
12.61	0.59	5.27	0.27	14.48	−0.13	2.74	0.19	1.87	7.06
3.33	13.57	8.09	4.59	4.8	18.22	4.07	12.39	−1.53	1.57
16.13	0.35	15.05	6.38	13.12	−0.8	5.6	6.54	5.23	−8.44
11.2	2.69	13.21	−0.24	−6.54	−5.75	−0.85	10.92	6.87	−5.72
1.14	18.45	1.72	10.32	−1.06	2.59	−0.28	−2.15	−1.69	6.95

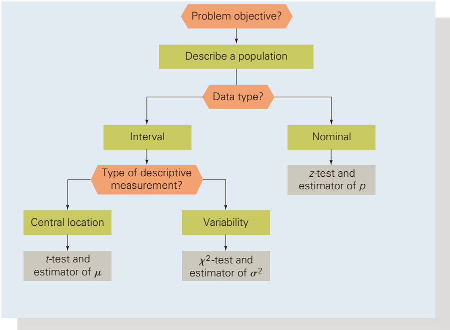
Can we conclude at the 5% significance level that directly purchased mutual funds outperform mutual funds bought through brokers?

SOLUTION:

IDENTIFY

To answer the question, we need to compare the population of returns from direct and the returns from broker-bought mutual funds. The data are obviously interval (we've recorded real numbers). This problem objective–data type combination tells us that the parameter to be tested is the difference between two means, $\mu_1 - \mu_2$. The hypothesis

*Source: D. Bergstresser, J. Chalmers, and P. Tufano, "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry."



Appendixes 13, 14, 15, 16, 17, and **19** reinforce this problem-solving approach and allow students to hone their skills.

Flowcharts, found within the appendixes, help students develop the logical process for choosing the correct technique, reinforce the learning process, and provide easy review material for students.

APPENDIX 14 Review of Chapters 12 to 14

The number of techniques introduced in Chapters 12 to 14 is up to 20. As we did in Appendix 13, we provide a table of the techniques with formulas and required conditions, a flowchart to help you identify the correct technique, and 25 exercises to give you practice in how to choose the appropriate method. The table and the flowchart have been amended to include the three analysis of variance techniques introduced in this chapter and the three multiple comparison methods.

TABLE **A14.1** **Summary of Statistical Techniques in Chapters 12 to 14**

t-test of μ
Estimator of μ (including estimator of $N\mu$)
χ^2 test of σ^2
Estimator of σ^2
z-test of p
Estimator of p (including estimator of Np)
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$
One-way analysis of variance (including multiple comparisons)
Two-way (randomized blocks) analysis of variance
Two-factor analysis of variance

Factors that Identify the t-Test and Estimator of μ_D

1. **Problem objective:** Compare two populations
2. **Data type:** Interval
3. **Descriptive measurement:** Central location
4. **Experimental design:** Matched pairs

Factors That Identify … boxes are found in each chapter after a technique or concept has been introduced. These boxes allow students to see a technique's essential requirements and give them a way to easily review their understanding. These essential requirements are revisited in the review chapters, where they are coupled with other concepts illustrated in flowcharts.

A Guide to Statistical Techniques, found on the inside front cover of the text, pulls everything together into one useful table that helps students identify which technique to perform based on the problem objective and data type. Here is part of the guide.

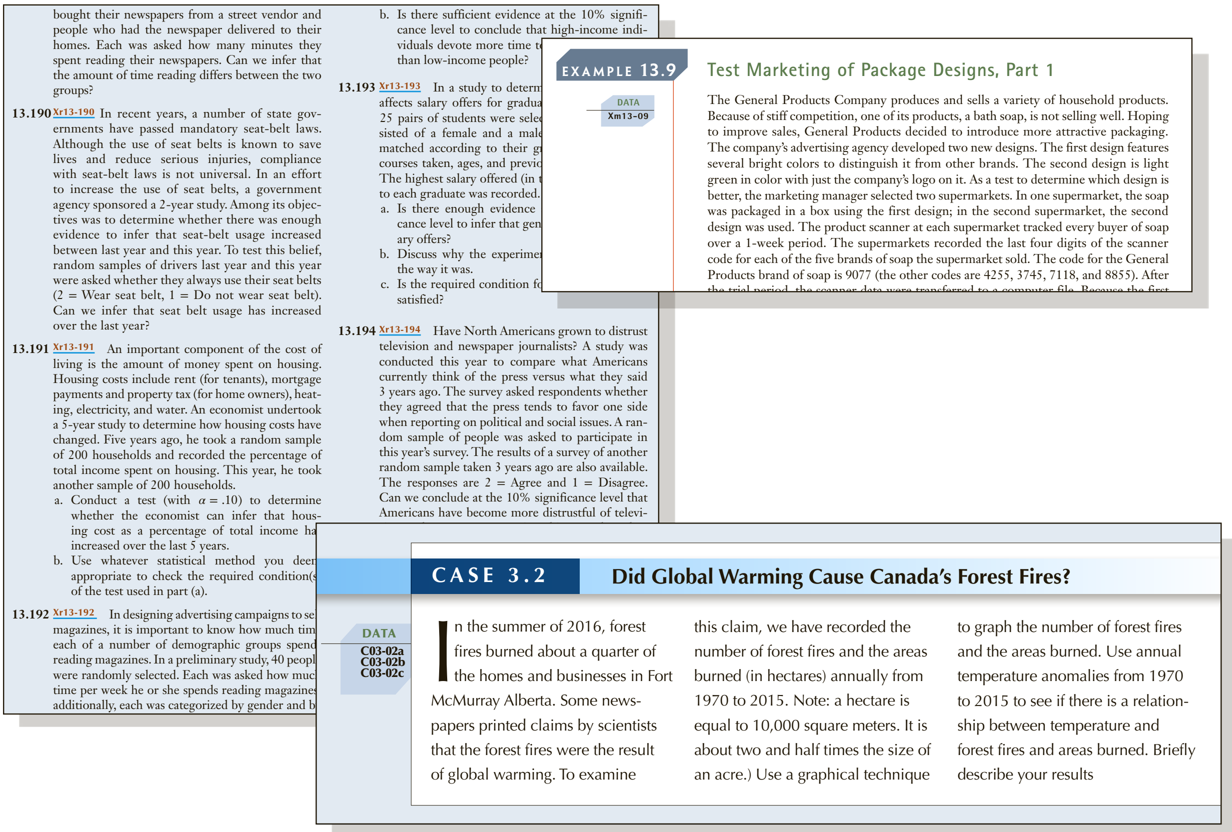
A GUIDE TO STATISTICAL TECHNIQUES

Problem Objectives

DATA TYPES	Describe a Population	Compare Two Populations	Compare Two or More Populations
Interval	Histogram **Section 3-1** Line chart **Section 3-2** Mean, median, and mode **Section 4-1** Range, variance, and standard deviation **Section 4-2** Percentiles and quartiles **Section 4-3** *t*-test and estimator of a mean **Section 12-1** Chi-squared test and estimator of a variance **Section 12-2**	Equal-variances *t*-test and estimator of the difference between two means: independent samples **Section 13-1** Unequal-variances *t*-test and estimator of the difference between two means: independent samples **Section 13-1** *t*-test and estimator of mean difference **Section 13-3** *F*-test and estimator of ratio of two variances **Section 13-4** Wilcoxon rank sum test **Section 19-1** Wilcoxon signed rank sum test **Section 19-2**	One-way analysis of variance **Section 14-1** LSD multiple comparison method **Section 14-2** Tukey's multiple comparison method **Section 14-2** Two-way analysis of variance **Section 14-4** Two-factor analysis of variance **Section 14-5** Kruskal-Wallis test **Section 19-3** Friedman test **Section 19-3**
Nominal	Frequency distribution **Section 2-2** Bar chart **Section 2-2** Pie chart **Section 2-2** *z*-test and estimator of a proportion **Section 12-3** Chi-squared goodness-of-fit test **Section 15-1**	*z*-test and estimator of the difference between two proportions **Section 13-5** Chi-squared test of a contingency table **Section 15-2**	Chi-squared test of a contingency table **Section 15-2**
Ordinal	Median **Section 4-1** Percentiles and quartiles **Section 4-3**	Wilcoxon rank sum test **Section 19-1** Sign test **Section 19-2**	Kruskal-Wallis test **Section 19-3** Friedman test **Section 19-3**

More Data Sets

A total of 1136 data sets available to be downloaded provide ample practice. These data sets contain real data, including stock market returns, climate change temperature anomalies and atmospheric carbon dioxide, baseball, basketball, football and hockey team payrolls, wins, and attendance.



bought their newspapers from a street vendor and people who had the newspaper delivered to their homes. Each was asked how many minutes they spent reading their newspapers. Can we infer that the amount of time reading differs between the two groups?

13.190 Xr13-190 In recent years, a number of state governments have passed mandatory seat-belt laws. Although the use of seat belts is known to save lives and reduce serious injuries, compliance with seat-belt laws is not universal. In an effort to increase the use of seat belts, a government agency sponsored a 2-year study. Among its objectives was to determine whether there was enough evidence to infer that seat-belt usage increased between last year and this year. To test this belief, random samples of drivers last year and this year were asked whether they always use their seat belts (2 = Wear seat belt, 1 = Do not wear seat belt). Can we infer that seat belt usage has increased over the last year?

13.191 Xr13-191 An important component of the cost of living is the amount of money spent on housing. Housing costs include rent (for tenants), mortgage payments and property tax (for home owners), heating, electricity, and water. An economist undertook a 5-year study to determine how housing costs have changed. Five years ago, he took a random sample of 200 households and recorded the percentage of total income spent on housing. This year, he took another sample of 200 households.
a. Conduct a test (with $\alpha = .10$) to determine whether the economist can infer that housing cost as a percentage of total income ha[s] increased over the last 5 years.
b. Use whatever statistical method you dee[m] appropriate to check the required condition(s) of the test used in part (a).

13.192 Xr13-192 In designing advertising campaigns to se[ll] magazines, it is important to know how much tim[e] each of a number of demographic groups spend[s] reading magazines. In a preliminary study, 40 peopl[e] were randomly selected. Each was asked how muc[h] time per week he or she spends reading magazines; additionally, each was categorized by gender and b[y]

b. Is there sufficient evidence at the 10% significance level to conclude that high-income individuals devote more time t[o] than low-income people?

13.193 Xr13-193 In a study to determ[ine] affects salary offers for gradua[tes] 25 pairs of students were sele[cted] sisted of a female and a male matched according to their g[...] courses taken, ages, and previ[...] The highest salary offered (in [...] to each graduate was recorded.
a. Is there enough evidence [...] cance level to infer that gen[...] ary offers?
b. Discuss why the experimen[...] the way it was.
c. Is the required condition f[...] satisfied?

13.194 Xr13-194 Have North Americans grown to distrust television and newspaper journalists? A study was conducted this year to compare what Americans currently think of the press versus what they said 3 years ago. The survey asked respondents whether they agreed that the press tends to favor one side when reporting on political and social issues. A random sample of people was asked to participate in this year's survey. The results of a survey of another random sample taken 3 years ago are also available. The responses are 2 = Agree and 1 = Disagree. Can we conclude at the 10% significance level that Americans have become more distrustful of televi-

EXAMPLE 13.9 Test Marketing of Package Designs, Part 1

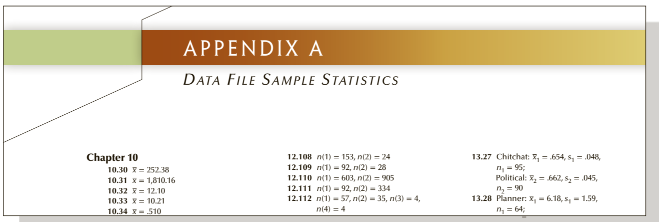
DATA Xm13-09

The General Products Company produces and sells a variety of household products. Because of stiff competition, one of its products, a bath soap, is not selling well. Hoping to improve sales, General Products decided to introduce more attractive packaging. The company's advertising agency developed two new designs. The first design features several bright colors to distinguish it from other brands. The second design is light green in color with just the company's logo on it. As a test to determine which design is better, the marketing manager selected two supermarkets. In one supermarket, the soap was packaged in a box using the first design; in the second supermarket, the second design was used. The product scanner at each supermarket tracked every buyer of soap over a 1-week period. The supermarkets recorded the last four digits of the scanner code for each of the five brands of soap the supermarket sold. The code for the General Products brand of soap is 9077 (the other codes are 4255, 3745, 7118, and 8855). After

CASE 3.2 Did Global Warming Cause Canada's Forest Fires?

DATA C03-02a C03-02b C03-02c

In the summer of 2016, forest fires burned about a quarter of the homes and businesses in Fort McMurray Alberta. Some newspapers printed claims by scientists that the forest fires were the result of global warming. To examine this claim, we have recorded the number of forest fires and the areas burned (in hectares) annually from 1970 to 2015. Note: a hectare is equal to 10,000 square meters. It is about two and half times the size of an acre.) Use a graphical technique to graph the number of forest fires and the areas burned. Use annual temperature anomalies from 1970 to 2015 to see if there is a relationship between temperature and forest fires and areas burned. Briefly describe your results



APPENDIX A

DATA FILE SAMPLE STATISTICS

Chapter 10
10.30 $\bar{x} = 252.38$
10.31 $\bar{x} = 1{,}810.16$
10.32 $\bar{x} = 12.10$
10.33 $\bar{x} = 10.21$
10.34 $\bar{x} = .510$

12.108 $n(1) = 153, n(2) = 24$
12.109 $n(1) = 92, n(2) = 28$
12.110 $n(1) = 603, n(2) = 905$
12.111 $n(1) = 92, n(2) = 334$
12.112 $n(1) = 57, n(2) = 35, n(3) = 4, n(4) = 4$

13.27 Chitchat: $\bar{x}_1 = .654, s_1 = .048, n_1 = 95$; Political: $\bar{x}_2 = .662, s_2 = .045, n_2 = 90$
13.28 Planner: $\bar{x}_1 = 6.18, s_1 = 1.59, n_1 = 64$;

Appendix A provides summary statistics for many of the exercises with large data sets. This feature offers unparalleled flexibility allowing students to solve most exercises by hand or by computer!

Real Data Sets

The data from the last eight General Social Surveys and the last four Surveys of Consumer Finances are included. These feature thousands of observations and dozens of selected variables. Solving more than 500 exercises associated with these surveys encourages students to uncover interesting aspects of the society. For example, students can determine the incomes, education, and working hours of people who are self-employed and compare them to people who work for someone else. They can see the effect of education on income, assets, investments, and net worth. Instructors can use the data to create their own examples and exercises.

Compute the Statistics

Once the correct technique has been identified, examples take students to the next level within the solution by asking them to compute the statistics.

EXCEL Data Analysis

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		*Finance*	*Marketing*
4	Mean	65,438	60,374
5	Variance	444,981,810	469,441,785
6	Observations	25	25
7	Pearson Correlation	0.9520	
8	Hypothesized Mean Difference	0	
9	df	24	
10	t Stat	3.81	
11	P(T<=t) one-tail	0.0004	
12	t Critical one-tail	1.7109	
13	P(T<=t) two-tail	0.0009	
14	t Critical two-tail	2.0639	

Excel prints the sample means, variances, and sample sizes for each sample (as well as the coefficient of correlation), which implies that the procedure uses these statistics. It doesn't. The technique is based on computing the paired differences from which the mean, variance, and sample size are determined. Excel should have printed these statistics.

INSTRUCTIONS

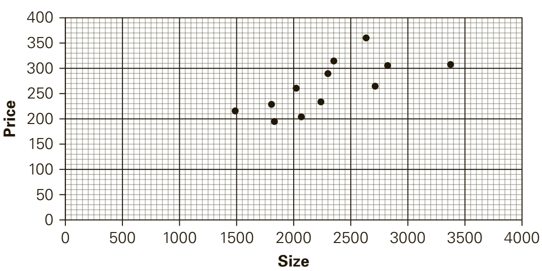
1. Type or import the data into two columns. (Open Xm13-05.)
2. Click **Data, Data Analysis,** and **t-Test: Paired Two-Sample for Means**.
3. Specify the **Variable 1 Range** (B1:B26) and the **Variable 2 Range** (C1:C26). Type the value of **the Hypothesized Mean Difference** (0) and specify a value for α (.05).

XLSTAT

	A	B	C	D	E	F
1	Hypothesized difference (D): 0					
2	Significance level (%): 5					
3						
4	Summary statistics					
5	Variable	Observations	Minimum	Maximum	Mean	Std. deviation
6	Finance	25	31,235	108,372	65,438	21,095
7	Marketing	25	29,682	93,205	60,374	21,667
8						
9	t-test for two paired samples / Upper-tailed test					
10	Difference	[illegible]				
11	t (Observed value)	3.81				
12	t (Critical value)	1.711				
13	DF	24				
14	p-value (one-tailed)	0.0004				
15	alpha	0.05				

Manual calculation of the problem is presented first in each "Compute" section of the examples.

Step-by-step instructions in the use of **Excel** and **XLSTAT** immediately follow the manual presentation. Instruction appears in the book with the printouts—there's no need to incur the extra expense of separate software manuals.



Ample use of graphics provides students many opportunities to see statistics in all its forms. In addition to manually presented figures throughout the text, Excel graphic outputs are given for students to compare to their own results.

Interpret the Results

INTERPRET

In the real world, it is not enough to know *how* to generate the statistics. To be truly effective, a business person must also know how to **interpret and articulate** the results. Furthermore, students need a framework to understand and apply statistics **within a realistic setting** by using realistic data in exercises, examples, and case studies.

Examples round out the final component of the identify–compute–interpret approach by asking students to interpret the results in the context of a business-related decision. This final step motivates and shows how statistics is used in everyday business situations.

An Applied Approach

With **Applications in …** sections and boxes, *Statistics for Management and Economics* now includes 42 **applications** (in finance, marketing, operations management, human resources, economics, and accounting) highlighting how statistics is used in those professions. For example, "Applications in Finance: Portfolio Diversification and Asset Allocation" shows how probability is used to help select stocks to minimize risk. Another optional section, "Applications in Marketing: Market Segmentation" demonstrates how to estimate the size of a market segment.

In addition to sections and boxes, **Applications in … exercises** can be found within the exercise sections to further reinforce the big picture.

APPLICATIONS in **OPERATIONS MANAGEMENT**

Michaeljung/Shutterstock.com

Quality

A critical aspect of production is quality. The quality of a final product is a function of the quality of the product's components. If the components don't fit, the product will not function as planned and likely cease functioning before its customers expect it to. For example, if a car door is not made to its specifications, it will not fit. As a result, the door will leak both water and air.

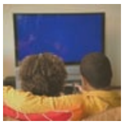
Operations managers attempt to maintain and improve the quality of products by ensuring that all components are made so that there is as little variation as possible. As you have already seen, statisticians measure variation by computing the variance.

Incidentally, an entire chapter (Chapter 21) is devoted to the topic of quality.

Nielsen Ratings

DATA Xm12-00*

Statistical techniques play a vital role in helping advertisers determine how many viewers watch the shows that they sponsor. Although several companies sample television viewers to determine what shows they watch, the best known is the A. C. Nielsen firm. The Nielsen ratings are based on a random sample of approximately 5,000 of the 115 million households in the United States with at least one television (in 2013). A meter attached to the televisions in the selected households keeps track of when the televisions are turned on and what channels they are tuned to. The data are sent to the Nielsen's computer every night from which Nielsen computes the rating and sponsors can determine the number of viewers and the potential value of any commercials. Of particular interest to advertisers are 18- to 49-year-olds, who are considered the most likely to buy advertised products. In 2013 there were 126.54 million Americans who were between 18 and 49 years old.

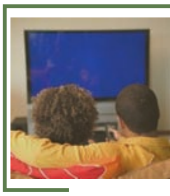


Jupiterimages

On page 415, we provide a solution to this problem.

Chapter-opening examples and solutions present compelling discussions of how the techniques and concepts introduced in that chapter are applied to real-world problems. These examples are then revisited with a solution as each chapter unfolds, applying the methodologies introduced in the chapter.

Nielsen Ratings: Solution



Jupiterimages

IDENTIFY

The problem objective is to describe the population of television shows watched by viewers across the country. The data are nominal. The combination of problem objective and data type make the parameter to be estimated the proportion of the entire population of 18- to 49-year-olds that watched *Big Bang Theory* (code = 2). The confidence interval estimator of the proportion is:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

COMPUTE

MANUALLY:

To solve manually, we count the number of 2's in the file. We find this value to be 275. Thus,

$$\hat{p} = \frac{x}{n} = \frac{275}{5{,}000} = .0550$$

The confidence level is $1 - \alpha = .95$. It follows that $\alpha = .05$, $\alpha/2 = .025$, and $z_{\alpha/2} = z_{.025} = 1.96$. The 95% confidence interval estimate of p is:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .0550 \pm 1.96\sqrt{\frac{(.0550)(1-.0550)}{5{,}000}} = .0550 \pm .0063$$

$$LCL = .0487 \qquad UCL = .0613$$

EXCEL Workbook

	A	B	C	D	E
1	z-Estimate of a Proportion				
2					
3	Sample proportion	0.055	Confidence Interval Estimate		
4	Sample size	5000	0.055	±	0.0063
5	Confidence level	0.95	Lower confidence limit		0.0487
6			Upper confidence limit		0.0613

INSTRUCTIONS

1. Type or import the data into one column. (Open Xm12-00.) In any empty cell, calculate the number of "successes" (=COUNTIF A1:A5001, 2). Divide that number by the sample size to obtain sample proportion.
2. Open the **Estimators Workbook** and click the ***z*-Estimate_Proportion** tab. Type or copy the sample proportion. Type the value of the sample size and the value of α.

XLSTAT

	A	B	C	D	E
1	Proportion: 0.055				
2	Sample size: 5000				
3					
4	95% confidence interval on the proportion (Wald):				
5	0.0487	0.0613			

CASE 12.5 Bias in Roulette Betting

DATA C12-05

The game of roulette consists of a wheel with 38 colored and numbered slots. The numbers are 1 to 36, 0 and 00. Half of the slots numbered 1 to 36 are red and the other half are black. The two "zeros" are green. The wheel is spun and an iron ball is rolled, which eventually comes to rest in one of the slots. Gamblers can make several different kinds of bets. Most players bet on one or more numbers or on a color (black or red). Here is the layout of the roulette betting table:

0	3	6	9	12	15	18	21	24	27	30	33	36
00	2	5	8	11	14	17	20	23	26	29	32	35
	1	4	7	10	13	16	19	22	25	28	31	34

Two statisticians recorded the bets on 904 spins. There were 21,731 bets.

Researchers wanted to use these data to examine *middle bias*, which is the tendency for guessers in multiple-choice exams to select the middle answers. For example, if there are five choices a, b, c, d, and e, guessers will tend to select answer c.

Most players stand on both sides of the betting table so that the middle numbers are 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, and 35.

a. If there is no middle bias, what proportion of the bets will be on 1 of the 12 middle numbers?

b. Conduct a test at the 5% significance level to determine whether middle bias exists.

c. The middle of the middle are the numbers 17 and 20. If there is no middle bias, what proportion of the bets will be either 17 or 20?

d. Test with a 5% significance level to determine whether middle of the middle bias exists.

Source: Maya Bar-Hillel and Ro'I Zultan, "We Sing the Praise of Good Displays: How Gamblers Bet in Casino Roulette," *Chance*, Volume 25, No. 2, 2012.

Many of the **examples, exercises, and cases are based on actual studies** performed by statisticians and published in journals, newspapers, and magazines, or presented at conferences. Many data files were recreated to produce the original results.

Chapter summaries briefly review the material and list important terms, symbols, and formulas.

CHAPTER SUMMARY

The inferential methods presented in this chapter address the problem of describing a single population. When the data are interval, the parameters of interest are the population mean μ and the population variance σ^2. The Student t-distribution is used to test and estimate the mean when the population standard deviation is unknown. The chi-squared distribution is used to make inferences about a population variance. When the data are nominal, the parameter to be tested and estimated is the population proportion p. The sample proportion follows an approximate normal distribution, which produces the test statistic and the interval estimator. We also discussed how to determine the sample size required to estimate a population proportion. We introduced market segmentation and described how statistical techniques presented in this chapter can be used to estimate the size of a segment.

IMPORTANT TERMS:

t-statistic 373
Student t-distribution 373
Robust 379
Chi-squared statistic 389

SYMBOLS:

Symbol	Pronounced	Represents
ν	nu	Degrees of freedom
χ^2	chi squared	Chi-squared statistic
$\hat{p}$	p hat	Sample proportion
$\tilde{p}$	p tilde	Wilson estimator

FORMULAS:

Test statistic for μ

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Confidence interval estimator of μ

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Confidence interval estimator of σ^2

$$\text{LCL} = \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

$$\text{UCL} = \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Test statistic for p

Instructor Resources

To access the instructor and student textbook resources, go to **www.cengage.com/login**, log in with your faculty account username and password, and use ISBN 9781337093453 to search for and add instructor resources to your account.

Student Learning Resources

To access student textbook resources, go to **www.cengagebrain.com** and search for this book by its title or its ISBN (9781337093453). You'll find Excel data sets, optional topics, and extra appendixes.

Acknowledgments

Although there is only one name on the cover of this book, the number of people who made contributions is large. I would like to acknowledge the work of all of them, with particular emphasis on the following: Paul Baum, California State University, Northridge, and John Lawrence, California State University, Fullerton, reviewed the page proofs. Their job was to find errors in presentation, arithmetic, and composition. The following individuals played important roles in the production of this book: Senior Product Manager Aaron Arnsparger, Content Developer Conor Allen, and Senior Content Project Manager Martha Conway. (For all remaining errors, place the blame where it belongs—on me.) Their advice and suggestions made my task considerably easier.

Fernando Rodriguez produced the test bank.

Trent Tucker, Wilfrid Laurier University, and Zvi Goldstein, California State University, Fullerton, each produced a set of PowerPoint slides.

The author extends thanks also to the survey participants and reviewers of the previous editions: Roger Bailey, Vanderbilt University; Paul Baum, California State University–Northridge; Nagraj Balakrishnan, Clemson University; Chen-Huei Chou, College of Charleston; Howard Clayton, Auburn University; Philip Cross, Georgetown University; Barry Cuffe, Wingate University; Ernest Demba, Washington University–St. Louis; Michael Douglas, Millersville University; Neal Duffy, State University of New York–Plattsburgh; John Dutton, North Carolina State University; Ehsan Elahi, University of Massachusetts–Boston; Erick Elder, University of Arkansas; Mohammed El-Saidi, Ferris State University; Grace Esimai, University of Texas–Arlington; Leila Farivar, The Ohio State University; Homi Fatemi, Santa Clara University; Abe Feinberg, California State University–Northridge; Samuel Graves, Boston College; Robert Gould, UCLA; Darren Grant, Sam Houston State University; Shane Griffith, Lee University; Paul Hagstrom, Hamilton College; John Hebert, Virginia Tech; James Hightower, California State University, Fullerton; Bo Honore, Princeton University; Ira Horowitz, University of Florida; Onisforos Iordanou, Hunter College; Torsten Jochem, University of Pittsburgh; Gordon Johnson, California State University–Northridge; Hilke Kayser, Hamilton College; Kenneth Klassen, California State University–Northridge; Roger Kleckner, Bowling Green State University–Firelands; Eylem Koca, Fairleigh Dickinson University; Harry Kypraios, Rollins College; John Lawrence, California State University–Fullerton; Tae H. Lee, University of California–Riverside; Dennis Lin, Pennsylvania State University; Jialu Liu, Allegheny College; Chung-Ping Loh, University of North Florida; Neal Long, Stetson University; Jayashree Mahajan, University of Florida; George Marcoulides, California State University–Fullerton;

Paul Mason, University of North Florida; Walter Mayer, University of Mississippi; John McDonald, Flinders University; Richard McGowan, Boston College; Richard McGrath, Bowling Green State University; Amy Miko, St. Francis College; Janis Miller, Clemson University; Glenn Milligan, Ohio State University; James Moran, Oregon State University; Robert G. Morris, University of Texas–Dallas; Patricia Mullins, University of Wisconsin; Adam Munson, University of Florida; David Murphy, Boston College; Kevin Murphy, Oakland University; Pin Ng, University of Illinois; Des Nicholls, Australian National University; Andrew Paizis, Queens College; David Pentico, Duquesne University; Ira Perelle, Mercy College; Nelson Perera, University of Wollongong; Bruce Pietrykowski, University of Michigan–Dearborn; Amy Puelz, Southern Methodist University; Lawrence Ries, University of Missouri; Colleen Quinn, Seneca College; Tony Quon, University of Ottawa; Madhu Rao, Bowling Green State University; Yaron Raviv, Claremont McKenna College; Jason Reed, Wayne State University; Phil Roth, Clemson University; Deb Rumsey, The Ohio State University; Farhad Saboori, Albright College; Don St. Jean, George Brown College; Hedayeh Samavati, Indiana–Purdue University; Sandy Shroeder, Ohio Northern University; Chris Silvia, University of Kansas; Jineshwar Singh, George Brown College; Natalia Smirnova, Queens College; Eric Sowey, University of New South Wales; Cyrus Stanier, Virginia Tech; Stan Stephenson, Southwest Texas State University; Gordon M. Stringer, University of Colorado–Colorado Springs; Arnold Stromberg, University of Kentucky; Pandu Tadikamalla, University of Pittsburgh; Patrick Thompson, University of Florida; Steve Thorpe, University of Northern Iowa; Sheldon Vernon, Houston Baptist University; John J. Wiorkowski, University of Texas–Dallas; and W. F. Younkin, University of Miami.

1

iStockphoto.com/leluconcepts

WHAT IS STATISTICS?

CHAPTER OUTLINE

Introduction

Statistics is a way to get information from data. That's it! Most of this textbook is devoted to describing how, when, and why managers and statistics practitioners* conduct statistical procedures. You may ask, "If that's all there is to statistics, why is this book (and most other statistics books) so large?" The answer is that students of applied statistics will be exposed to different kinds of information and data. We demonstrate some of these with a case and two examples that are featured later in this book.

The first may be of particular interest to you.

---

*The term *statistician* is used to describe so many different kinds of occupations that it has ceased to have any meaning. It is used, for example, to describe a person who calculates baseball statistics as well as an individual educated in statistical principles. We will describe the former as a *statistics practitioner* and the

(*continued*)

EXAMPLE 3.3

Business Statistics Marks (See Chapter 3)

A student enrolled in a business program is attending his first class of the required statistics course. The student is somewhat apprehensive because he believes the myth that the course is difficult. To alleviate his anxiety, the student asks the professor about last year's marks. Because this professor is friendly and helpful, like all other statistics professors, he obliges the student and provides a list of the final marks, which are composed of term work plus the final exam. What information can the student obtain from the list?

This is a typical statistics problem. The student has the data (marks) and needs to apply statistical techniques to get the information he requires. This is a function of **descriptive statistics**.

Descriptive Statistics

Descriptive statistics deals with methods of organizing, summarizing, and presenting data in a convenient and informative way. One form of descriptive statistics uses graphical techniques that allow statistics practitioners to present data in ways that make it easy for the reader to extract useful information. In Chapters 2 and 3 we will present a variety of graphical methods.

Another form of descriptive statistics uses numerical techniques to summarize data. One such method that you have already used frequently calculates the average or mean. In the same way that you calculate the average age of the employees of a company, we can compute the mean mark of last year's statistics course. Chapter 4 introduces several numerical statistical measures that describe different features of the data.

The actual technique we use depends on what specific information we would like to extract. In this example, we can see at least three important pieces of information. The first is the "typical" mark. We call this a *measure of central location*. The average is one such measure. In Chapter 4, we will introduce another useful measure of central location, the median. Suppose the student was told that the average mark last year was 67. Is this enough information to reduce his anxiety? The student would likely respond "No" because he would like to know whether most of the marks were close to 67 or were scattered far below and above the average. He needs a *measure of variability*. The simplest such measure is the *range*, which is calculated by subtracting the smallest number from the largest. Suppose the largest mark is 96 and the smallest is 24. Unfortunately, this provides little information since it is based on only two marks. We need other measures—these will be introduced in Chapter 4. Moreover, the student must determine more about the marks. In particular, he needs to know how the marks are distributed between 24 and 96. The best way to do this is to use a graphical technique, the histogram, which will be introduced in Chapter 3.

---

latter as a *statistician*. A statistics practitioner is a person who uses statistical techniques properly. Examples of statistics practitioners include the following:

1. a financial analyst who develops stock portfolios based on historical rates of return;
2. an economist who uses statistical models to help explain and predict variables such as inflation rate, unemployment rate, and changes in the gross domestic product; and
3. a market researcher who surveys consumers and converts the responses into useful information.

Our goal in this book is to convert you into one such capable individual.

The term *statistician* refers to an individual who works with the mathematics of statistics. His or her work involves research that develops techniques and concepts, which in the future may help the statistics practitioner. Statisticians are also statistics practitioners, frequently conducting empirical research and consulting. If you're taking a statistics course, your instructor is probably a statistician.

Case 12.1 Pepsi's Exclusivity Agreement with a University (see Chapter 12) In the last few years, colleges and universities have signed exclusivity agreements with a variety of private companies. These agreements bind the university to sell these companies' products exclusively on the campus. Many of the agreements involve food and beverage firms.

A large university with a total enrollment of about 50,000 students has offered Pepsi-Cola an exclusivity agreement that would give Pepsi exclusive rights to sell its products at all university facilities for the next year with an option for future years. In return, the university would receive 35% of the on-campus revenues and an additional lump sum of $200,000 per year. Pepsi has been given 2 weeks to respond.

The management at Pepsi quickly reviews what it knows. The market for soft drinks is measured in terms of 12-ounce cans. Pepsi currently sells an average of 22,000 cans per week over the 40 weeks of the year that the university operates. The cans sell for an average of one dollar each. The costs, including labor, total 30 cents per can. Pepsi is unsure of its market share but suspects it is considerably less than 50%. A quick analysis reveals that if its current market share were 25%, then, with an exclusivity agreement, Pepsi would sell 88,000 (22,000 is 25% of 88,000) cans per week or 3,520,000 cans per year. The gross revenue would be computed as follows[†]:

$$\text{Gross revenue} = 3{,}520{,}000 \times \$1.00/\text{can} = \$3{,}520{,}000$$

This figure must be multiplied by 65% because the university would rake in 35% of the gross. Thus,

$$\text{Gross revenue after deducting 35\% university take} \\ = 65\% \times \$3{,}520{,}000 = \$2{,}288{,}000$$

The total cost of 30 cents per can (or $1,056,000) and the annual payment to the university of $200,000 are subtracted to obtain the net profit:

$$\text{Net profit} = \$2{,}288{,}000 - \$1{,}056{,}000 - \$200{,}000 = \$1{,}032{,}000$$

Pepsi's current annual profit is

$$40 \text{ weeks} \times 22{,}000 \text{ cans/week} \times \$.70 = \$616{,}000$$

If the current market share is 25%, the potential gain from the agreement is

$$\$1{,}032{,}000 - \$616{,}000 = \$416{,}000$$

The only problem with this analysis is that Pepsi does not know how many soft drinks are sold weekly at the university. Coke is not likely to supply Pepsi with information about its sales, which together with Pepsi's line of products constitute virtually the entire market.

Pepsi assigned a recent university graduate to survey the university's students to supply the missing information. Accordingly, she organizes a survey that asks 500 students to keep track of the number of soft drinks they purchase in the next 7 days. The responses are stored in a file C12-01 available to be downloaded.

Inferential Statistics

The information we would like to acquire in Case 12.1 is an estimate of annual profits from the exclusivity agreement. The data are the numbers of cans of soft drinks consumed in 7 days by the 500 students in the sample. We can use descriptive techniques to

[†]We have created an Excel spreadsheet that does the calculations for this case. See Appendix 1 for instructions on how to download this spreadsheet from Cengage's website plus hundreds of data sets and much more.

learn more about the data. In this case, however, we are not so much interested in what the 500 students are reporting as in knowing the mean number of soft drinks consumed by all 50,000 students on campus. To accomplish this goal we need another branch of statistics: **inferential statistics**.

Inferential statistics is a body of methods used to draw conclusions or inferences about characteristics of populations based on sample data. The population in question in this case is the university's 50,000 students. The characteristic of interest is the soft drink consumption of this population. The cost of interviewing each student in the population would be prohibitive and extremely time consuming. Statistical techniques make such endeavors unnecessary. Instead, we can sample a much smaller number of students (the sample size is 500) and infer from the data the number of soft drinks consumed by all 50,000 students. We can then estimate annual profits for Pepsi.

EXAMPLE 12.5

Exit Polls (See Chapter 12)

When an election for political office takes place, the television networks cancel regular programming to provide election coverage. After the ballots are counted, the results are reported. However, for important offices such as president or senator in large states, the networks actively compete to see which one will be the first to predict a winner. This is done through **exit polls** in which a random sample of voters who exit the polling booth are asked for whom they voted. From the data, the sample proportion of voters supporting the candidates is computed. A statistical technique is applied to determine whether there is enough evidence to infer that the leading candidate will garner enough votes to win. Suppose that the exit poll results from the state of Florida during the year 2000 elections were recorded. Although several candidates were running for president, the exit pollsters recorded only the votes of the two candidates who had any chance of winning: Republican George W. Bush and Democrat Albert Gore. The results (765 people who voted for either Bush or Gore) were stored in file Xm12-05. The network analysts would like to know whether they can conclude that George W. Bush will win the state of Florida.

Example 12.5 describes a common application of statistical inference. The population the television networks wanted to make inferences about is the approximately 5 million Floridians who voted for Bush or Gore for president. The sample consisted of the 765 people randomly selected by the polling company who voted for either of the two main candidates. The characteristic of the population that we would like to know is the proportion of the Florida total electorate that voted for Bush. Specifically, we would like to know whether more than 50% of the electorate voted for Bush (counting only those who voted for either the Republican or Democratic candidate). It must be made clear that we cannot predict the outcome with 100% certainty because we will not ask all 5 million actual voters for whom they voted. This is a fact that statistics practitioners and even students of statistics must understand. A sample that is only a small fraction of the size of the population can lead to correct inferences only a certain percentage of the time. You will find that statistics practitioners can control that fraction and usually set it between 90% and 99%.

Incidentally, on the night of the U.S. election in November 2000, the networks goofed badly. Using exit polls as well as the results of previous elections, all four networks concluded at about 8 P.M. that Al Gore would win Florida. Shortly after 10 P.M., with a large percentage of the actual vote having been counted, the networks reversed course and declared that George W. Bush would win the state. By 2 A.M., another verdict was declared: The result was too close to call. Since then, this experience has likely been used by statistics instructors when teaching how *not* to use statistics.

Notice that, contrary to what you probably believed, data are not necessarily numbers. The marks in Example 3.3 and the number of soft drinks consumed in a week in Case 12.1, of course, are numbers; however, the votes in Example 12.5 are not. In Chapter 2, we will discuss the different types of data you will encounter in statistical applications and how to deal with them.

1-1 Key Statistical Concepts

Statistical inference problems involve three key concepts: the population, the sample, and the statistical inference. We now discuss each of these concepts in more detail.

1-1a Population

A **population** is the group of all items of interest to a statistics practitioner. It is frequently very large and may, in fact, be infinitely large. In the language of statistics, *population* does not necessarily refer to a group of people. It may, for example, refer to the population of ball bearings produced at a large plant. In Case 12.1, the population of interest consists of the 50,000 students on campus. In Example 12.5, the population consists of the Floridians who voted for Bush or Gore.

A descriptive measure of a population is called a **parameter**. The parameter of interest in Case 12.1 is the mean number of soft drinks consumed by all the students at the university. The parameter in Example 12.5 is the proportion of the 5 million Florida voters who voted for Bush. In most applications of inferential statistics, the parameter represents the information we need.

1-1b Sample

A **sample** is a set of data drawn from the studied population. A descriptive measure of a sample is called a **statistic**. We use statistics to make inferences about parameters. In Case 12.1, the statistic we would compute is the mean number of soft drinks consumed in the last week by the 500 students in the sample. We would then use the sample mean to infer the value of the population mean, which is the parameter of interest in this problem. In Example 12.5, we compute the proportion of the sample of 765 Floridians who voted for Bush. The sample statistic is then used to make inferences about the population of all 5 million votes—that is, we predict the election results even before the actual count.

1-1c Statistical Inference

Statistical inference is the process of making an estimate, prediction, or decision about a population based on sample data. Because populations are almost always very large, investigating each member of the population would be impractical and expensive. It is far easier and cheaper to take a sample from the population of interest and draw conclusions or make estimates about the population on the basis of information provided by the sample. However, such conclusions and estimates are not always going to be correct. For this reason, we build into the statistical inference a measure of reliability. There are two such measures: the **confidence level** and the **significance level**. The *confidence level* is the proportion of times that an estimating procedure will be correct. For example, in Case 12.1, we will produce an estimate of the average number of soft drinks to be consumed by all 50,000 students that has a confidence level of 95%. In other words, estimates based on this form of statistical inference will be correct 95% of the time. When the purpose of the statistical inference is to draw a conclusion about a population, the *significance level* measures how frequently the conclusion will be wrong. For example, suppose that, as a result of the analysis in Example 12.5, we conclude that more than 50% of the electorate will vote for George W. Bush, and thus he will win the state of Florida. A 5% significance level means that samples that lead us to conclude that Bush wins the election will be wrong 5% of the time.

1-2 Statistical Applications in Business

An important function of statistics courses in business and economics programs is to demonstrate that statistical analysis plays an important role in virtually all aspects of business and economics. We intend to do so through examples, exercises, and cases. However, we assume that most students taking their first statistics course have not taken courses in most of the other subjects in management programs. To understand fully how statistics is used in these and other subjects, it is necessary to know something about them. To provide sufficient background to understand the statistical application, we introduce applications in accounting, economics, finance, human resources management, marketing, and operations management. We provide readers with some background of these applications by describing their functions in two ways.

1-2a Application Sections and Subsections

We feature five sections that describe statistical applications in the functional areas of business. In Section 4-5, we discuss an application in finance, the market model, which introduces an important concept in investing. Section 7-3 describes another application in finance that describes a financial analyst's use of probability and statistics to construct portfolios that decrease risk. Section 12-4 is an application in marketing, market segmentation. In Section 14-6, we present an application in operations management, finding and reducing variation. In Section 18-3, we provide an application in human resources, pay equity. A subsection in Section 6-4 presents an application in medical testing (useful in the medical insurance industry).

1-2b Application Boxes

For other topics that require less-detailed description, we provide application boxes with a relatively brief description of the background followed by examples or exercises. These boxes are scattered throughout the book. For example, in Section 4-1, we discuss the geometric mean and why it is used instead of the arithmetic mean to measure variables that are rates of change.

1-3 Large Real Data Sets

The author believes that you learn statistics by doing statistics. For their lives after college and university, we expect graduates to have access to large amounts of real data that must be summarized to acquire the information needed to make decisions. We include the data from two sources: the General Social Survey (GSS) and the Survey of Consumer Finances (SCF). We have scattered examples, exercises, and cases for these surveys throughout the book.

1-3a General Social Survey

Since 1972, the GSS has been tracking American attitudes on a wide variety of topics. With the exception of the U.S. Census, the GSS is the most frequently used source of information about American society. The surveys now conducted every second year measure hundreds of variables and thousands of observations. The data for the eight most recent surveys are stored in files GSS2000, GSS2002, GSS2004, GSS2006, GSS2008, GSS2010, GSS2012, and GSS2014. The sample sizes are 2,817, 2,765, 2,812, 4,510, 2,023, 2,044, 1,974, and 2,538, respectively. We have removed the missing data codes representing "No answer," "Don't know," and so on and replaced them with blanks. Be aware that Excel and XLSTAT have different ways of dealing with blanks.

A list of all the variables and their definitions is available as an online appendix.

1-3b Survey of Consumer Finances

The SCF is conducted every 3 years to provide detailed information on the finances of U.S. households. The study is sponsored by the Federal Reserve Board in cooperation with the Department of the Treasury. Since 1992, data have been collected by the National Opinion Research Center (NORC) at the University of Chicago. The data for the four most recent surveys are stored in folders SCF2004, SCF2007, SCF2010, and SCF2013. The sample sizes are 4,519, 4,417, 6,482, and 6,015, respectively. Because the samples are so large and the range of some of the variables so wide, there are problems summarizing and describing the data. To solve the problem, we have created subsamples based on percentiles of the net worth of the households being sampled. Here is a list of the subsamples.

All: Includes all observations

B1: Lowest 1%

Poor (P): 1%–5%

Working poor (WP): 5%–20%

Lower Middle Class (LMC): 20%–40%

Middle Class (MC): 40%–60%

Upper Middle Class (UMC): 60%–80%

Upper Class (UC): 80%–90%

Wealthy (W): 90%–95%

Super Rich (SR): 95%–99%

T1: Top 1%

A complete list of the variables and their definitions is available as an online appendix.

1-4 Statistics and the Computer

In virtually all applications of statistics, the statistics practitioner must deal with large amounts of data. For example, Case 12.1 (Pepsi-Cola) involves 500 observations. To estimate annual profits, the statistics practitioner would have to perform computations on the data. Although the calculations do not require any great mathematical skill, the sheer amount of arithmetic makes this aspect of the statistical method time consuming and tedious. Fortunately, numerous commercially prepared computer programs are available to perform the arithmetic. We have chosen to use Microsoft Excel in the belief that virtually all university graduates use it now and will in the future.

1-4a Excel

Excel can perform statistical procedures in several ways.

1. **Statistical** (which includes probability) and other functions f_x**:** We use some of these functions to help create graphical techniques in Chapter 2, calculate statistics in Chapters 3 and 4, and to compute probabilities in Chapters 7 and 8.
2. **Spreadsheets:** We use statistical functions to create spreadsheets that calculate statistical inference methods in Chapters 10–16. These can be downloaded from Cengage's website. Additionally, the spreadsheets can be used to conduct what–if analyses. The rationale for their use is described in subsection 1-4f.
3. **Analysis ToolPak:** This group of procedures comes with every version of Excel. The techniques are accessed by clicking Data and Data Analysis. One of its drawbacks is that it does not offer a complete set of the statistical techniques we introduce in this book.
4. **XLSTAT** is a commercially created add-in that can be loaded onto your computer to enable you to use Excel for almost all statistical procedures introduced in this book. XLSTAT can be downloaded from Cengage's website.

1-4b Data Analysis Plus

We have offered Data Analysis Plus in the last seven editions of this book. Unfortunately, we have encountered problems with some of the combinations of Excel versions and operating systems. As a result it is no longer possible to offer Data Analysis Plus as a universal tool for all Excel users of this book. Appendix 1 lists the combinations that do work. Data Analysis Plus can be downloaded from the author's website. Printouts and instructions are available as appendixes that can be downloaded from the Cengage's website.

1-4c Minitab

Like Data Analysis Plus, Minitab has been used in many editions of this book. Because of decreasing demand, we have decided to discontinue Minitab use. However, printouts and instructions for Minitab 17 are available as online appendixes.

1-4d File names and Notation

A large proportion of the examples, exercises, and cases feature large data sets. These are denoted with the file name next to the exercise number. The data sets associated with examples are denoted as Xm. To illustrate, the data for Example 2.2 are stored in file Xm02-02 in the Chapter 2 folder. The data for exercises and cases are stored in files prefixed by Xr and C, respectively. The prefix GSS designates data for the General Social Surveys. The data for the Surveys of Consumer Finances data are stored in folders SCF2004, SCF2007, SCF2010, and SCF2013.

In many real applications of statistics, additional data are collected. For instance, in Example 12.5, the pollster often records the voter's gender and asks for other information including race, religion, education, and income. Many other data sets are similarly constructed. In later chapters, we will return to these files and require other statistical techniques to extract the needed information. (Files that contain additional data are denoted by an asterisk on the file name.)

1-4e Our Approach

The approach we prefer to take is to minimize the time spent on manual computations and to focus instead on selecting the appropriate method for dealing with a problem and on interpreting the output after the computer has performed the necessary computations. In this way, we hope to demonstrate that statistics can be as interesting and as practical as any other subject in your curriculum.

1-4f Excel Spreadsheets

Books written for statistics courses taken by mathematics or statistics majors are considerably different from this one. It is not surprising that such courses feature mathematical proofs of theorems and derivations of most procedures. When the material is covered in this way, the underlying concepts that support statistical inference are exposed and relatively easy to see. However, this book was created for an applied course in business and economics statistics. Consequently, we do not address directly the mathematical principles of statistics. However, as we pointed out previously, one of the most important functions of statistics practitioners is to properly interpret statistical results, whether produced manually or by computer. And, to correctly interpret statistics, students require an understanding of the principles of statistics.

To help students understand the basic foundation, we offer readers Excel spreadsheets that allow for *what–if* analyses. By changing some of the input value, students can see for themselves how statistics works. (The term is derived from *what* happens to the statistics *if* I change this value?)

Chapter Summary

IMPORTANT TERMS:

Descriptive statistics 2
Inferential statistics 4
Exit polls 4
Population 5
Parameter 5
Sample 5
Statistic 5
Statistical inference 5
Confidence level 5
Significance level 5