Third edition

Statistics for Business and Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams
Jim Freeman
Eddie Shoesmith

Third edition

Statistics for Business and Economics

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams
Jim Freeman
Eddie Shoesmith

CENGAGE Learning

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

Statistics for Business and Economics, Third Edition
David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jim Freeman and Eddie Shoesmith

Publishing Director: Linden Harris
Publisher: Andrew Ashwin
Development Editor: Felix Rowe
Production Editor: Beverley Copland
Manufacturing Buyer: Elaine Willis
Marketing Manager: Vicky Fielding
Typesetter: Integra Software Services Pvt. Ltd.
Cover design: Adam Renvoize

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN: 978-1-4080-7223-3

Cengage Learning EMEA
Cheriton House, North Way, Andover, Hampshire, SP10 5BE, United Kingdom

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

For your lifelong learning solutions, visit **www.cengage.co.uk**

Purchase your next print book, e-book or e-chapter at **www.cengagebrain.com**

Printed in China by RR Donnelley
1 2 3 4 5 6 7 8 9 10 – 16 15 14

BRIEF CONTENTS

Book contents

Online contents

CONTENTS

Online contents

DEDICATION

'To the memory of my grandparents, Lizzie and Halsey'

JIM FREEMAN

'To all my family, past, present and future'

EDDIE SHOESMITH

PREFACE

The purpose of *Statistics for Business and Economics* is to give students, primarily those in the fields of business, management and economics, a conceptual introduction to the field of statistics and its many applications. The text is applications oriented and written with the needs of the non-mathematician in mind. The mathematical prerequisite is knowledge of algebra.

Applications of data analysis and statistical methodology are an integral part of the organization and presentation of the material in the text. The discussion and development of each technique are presented in an application setting, with the statistical results providing insights to problem solution and decision-making.

Although the book is applications oriented, care has been taken to provide sound methodological development and to use notation that is generally accepted for the topic being covered. Hence, students will find that this text provides good preparation for the study of more advanced statistical material. A revised and updated bibliography to guide further study is included as an appendix.

The online platform introduces the student to the software packages MINITAB 16, SPSS 21 and Microsoft® Office EXCEL 2010, and emphasizes the role of computer software in the application of statistical analysis. MINITAB and SPSS are illustrated as they are two of the leading statistical software packages for both education and statistical practice. EXCEL is not a statistical software package, but the wide availability and use of EXCEL makes it important for students to understand the statistical capabilities of this package. MINITAB, SPSS and EXCEL procedures are provided on the dedicated online platform so that instructors have the flexibility of using as much computer emphasis as desired for the course.

THE EMEA EDITION

This is the 3rd EMEA edition of *Statistics for Business and Economics*. It is based on the 2nd EMEA edition and the 11th United States (US) edition. The US editions have a distinguished history and deservedly high reputation for clarity and soundness of approach, and we maintained the presentation style and readability of those editions in preparing the international edition. We have replaced many of the US-based examples, case studies and exercises with equally interesting and appropriate ones sourced from a wider geographical base, particularly the UK, Ireland, continental Europe, South Africa and the Middle East. We have also streamlined the book by moving four non-mandatory chapters, the software section and exercise answers to the associated online platform. Other notable changes in this 3rd EMEA edition are summarized here.

CHANGES IN THE 3RD EMEA EDITION

- **Self-test exercises** Certain exercises are identified as self-test exercises. Completely worked-out solutions for those exercises are provided on the online platform that accompanies the text. Students can attempt the self-test exercises and immediately check the solution to evaluate their understanding of the concepts presented in the chapter.

- **Other content revisions** The following additional content revisions appear in the new edition:
  - New examples of times series data are provided in Chapter 1.
  - Chapter 9 contains a revised introduction to hypothesis testing, with a better set of guidelines for identifying the null and alternative hypotheses.
  - Chapter 13 makes much more explicit the linkage between Analysis of Variance and experimental design.
  - Chapter 17 now includes coverage of the popular Holt's linear exponential smoothing methodology.
  - The treatment of non-parametric methods in Chapter 18 has been revised and updated.
  - Chapter 19 on index numbers (on the online platform) has been updated with current index numbers.
  - A number of case problems have been added or updated. These are in the chapters on Descriptive Statistics, Discrete Probability Distributions, Inferences about Population Variances, Tests of Goodness of Fit and Independence, Simple Linear Regression, Multiple Regression, Regression Analysis: Model Building, Non-Parametric Methods, Index Numbers and Decision Analysis. These case problems provide students with the opportunity to analyze somewhat larger data sets and prepare managerial reports based on the results of the analysis.
  - Each chapter begins with a Statistics in Practice article that describes an application of the statistical methodology to be covered in the chapter. New to this edition are Statistics in Practice articles for Chapters 2, 9, 10 and 11, with several other articles substantially updated and revised for this new edition.
  - New examples and exercises have been added throughout the book, based on real data and recent reference sources of statistical information. We believe that the use of real data helps generate more student interest in the material and enables the student to learn about both the statistical methodology and its application.
  - To accompany the new exercises and examples, data files are available on the online platform. The data sets are available in MINITAB, SPSS and EXCEL formats. Data set logos are used in the text to identify the data sets that are available on the online platform. Data sets for all case problems as well as data sets for larger exercises are included.
- **Software sections** In the 3rd EMEA edition, we have updated the software sections to provide step-by-step instructions for the latest versions of the software packages: MINITAB 16, SPSS 21 and Microsoft® Office EXCEL 2010. The software sections have been relocated to the online platform.

ACKNOWLEDGEMENTS

The authors and publisher acknowledge the contribution of the following reviewers throughout the three editions of this textbook:

- John R. Calvert – Loughborough University (UK)
- Naomi Feldman – Ben-Gurion University of the Negev (Israel)
- Luc Hens – Vesalius College (Belgium)
- Martyn Jarvis – University of Glamorgan (UK)
- Khalid M Kisswani – Gulf University for Science & Technology (Kuwait)
- Alan Matthews – Trinity College Dublin (Ireland)
- Suzanne McCallum – Glasgow University (UK)
- Chris Muller – University of Stellenbosch (South Africa)
- Surette Oosthuizen – University of Stellenbosch (South Africa)
- Karim Sadrieh – Otto von Guericke University Magdeburg (Germany)
- Mark Stevenson – Lancaster University (UK)
- Dave Worthington – Lancaster University (UK)
- Zhan Pang – Lancaster University (UK)

ABOUT THE AUTHORS

Jim Freeman is Senior Lecturer in Statistics and Operational Research at Manchester Business School (MBS), United Kingdom. He was born in Tewkesbury, Gloucestershire. After taking a first degree in pure mathematics at UCW Aberystwyth, he went on to receive MSc and PhD degrees in Applied Statistics from Bath and Salford universities respectively. In 1992/3 he was Visiting Professor at the University of Alberta. Before joining MBS, he was Statistician at the Distributive Industries Training Board – and prior to that – the Universities Central Council on Admissions. He has taught undergraduate and postgraduate courses in business statistics and operational research courses to students from a wide range of management and engineering backgrounds. For many years he was also responsible for providing introductory statistics courses to staff and research students at the University of Manchester's Staff Teaching Workshop. Through his gaming and simulation interests he has been involved in a significant number of external consultancy projects. In July 2008 he was appointed Editor of the Operational Research Society's *OR Insight* journal.

Eddie Shoesmith was formerly Senior Lecturer in Statistics and Programme Director for undergraduate business and management programmes in the School of Business, University of Buckingham, UK. He was born in Barnsley, Yorkshire. He was awarded an MA (Natural Sciences) at the University of Cambridge, and a BPhil (Economics and Statistics) at the University of York. Prior to taking an academic post at Buckingham, he worked for the UK Government Statistical Service, in the Cabinet Office, for the London Borough of Hammersmith and for the London Borough of Haringey. At Buckingham, before joining the School of Business, he held posts as Dean of Sciences and Head of Psychology. He has taught introductory and intermediate-level applied statistics courses to undergraduate and postgraduate student groups in a wide range of disciplines: business and management, economics, accounting, psychology, biology and social sciences. He has also taught statistics to social and political sciences undergraduates at the University of Cambridge.

David R. Anderson is Professor of Quantitative Analysis in the College of Business Administration at the University of Cincinnati. Born in Grand Forks, North Dakota, he earned his BS, MS and PhD degrees from Purdue University. Professor Anderson has served as Head of the Department of Quantitative Analysis and Operations Management and as Associate Dean of the College of Business Administration. In addition, he was the coordinator of the college's first executive programme. In addition to teaching introductory statistics for business students, Dr Anderson has taught graduate-level courses in regression analysis, multivariate analysis and management science. He also has taught statistical courses at the Department of Labor in Washington, DC. Professor Anderson has been honoured with nominations and awards for excellence in teaching and excellence in service to student organizations. He has co-authored ten textbooks related to decision sciences and actively consults with businesses in the areas of sampling and statistical methods.

Dennis J. Sweeney is Professor of Quantitative Analysis and founder of the Center for Productivity Improvement at the University of Cincinnati. Born in Des Moines, Iowa, he earned BS and BA degrees from Drake University, graduating *summa cum laude*. He received his MBA and DBA degrees from Indiana University, where he was an NDEA Fellow. Dr Sweeney has worked in the management science

group at Procter & Gamble and has been a visiting professor at Duke University. Professor Sweeney served five years as Head of the Department of Quantitative Analysis and four years as Associate Dean of the College of Business Administration at the University of Cincinnati.

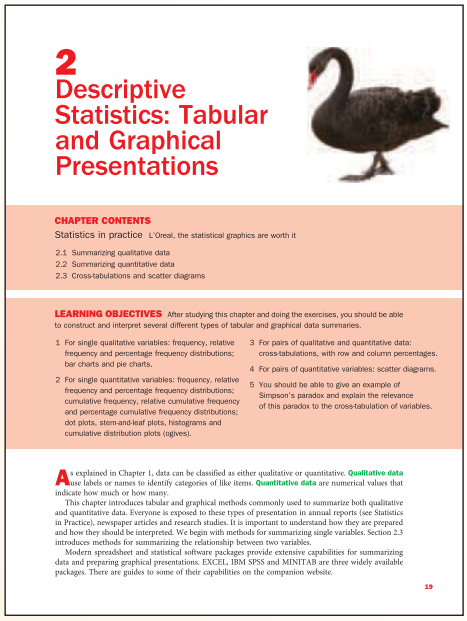
He has published more than 30 articles in the area of management science and statistics. The National Science Foundation, IBM, Procter & Gamble, Federated Department Stores, Kroger and Cincinnati Gas & Electric have funded his research, which has been published in *Management Science, Operations Research, Mathematical Programming, Decision Sciences* and other journals. Professor Sweeney has co-authored ten textbooks in the areas of statistics, management science, linear programming and production and operations management.

Thomas A. Williams is Professor of Management Science in the College of Business at Rochester Institute of Technology (RIT). Born in Elmira, New York, he earned his BS degree at Clarkson University. He completed his graduate work at Rensselaer Polytechnic Institute, where he received his MS and PhD degrees.

Before joining the College of Business at RIT, Professor Williams served for seven years as a faculty member in the College of Business Administration at the University of Cincinnati, where he developed the first undergraduate programme in Information Systems. At RIT he was the first chair of the Decision Sciences Department.

Professor Williams is the co-author of 11 textbooks in the areas of management science, statistics, production and operations management and mathematics. He has been a consultant for numerous *Fortune* 500 companies in areas ranging from the use of elementary data analysis to the development of large-scale regression models.

WALK-THROUGH TOUR



2
Descriptive Statistics: Tabular and Graphical Presentations

CHAPTER CONTENTS

Statistics in practice L'Oreal, the statistical graphics are worth it

2.1 Summarizing qualitative data
2.2 Summarizing quantitative data
2.3 Cross-tabulations and scatter diagrams

LEARNING OBJECTIVES After studying this chapter and doing the exercises, you should be able to construct and interpret several different types of tabular and graphical data summaries.

1 For single qualitative variables: frequency, relative frequency and percentage frequency distributions; bar charts and pie charts.
2 For single quantitative variables: frequency, relative frequency and percentage frequency distributions; cumulative frequency, relative cumulative frequency and percentage cumulative frequency distributions; dot plots, stem-and-leaf plots, histograms and cumulative distribution plots (ogives).
3 For pairs of qualitative and quantitative data: cross-tabulations, with row and column percentages.
4 For pairs of quantitative variables: scatter diagrams.
5 You should be able to give an example of Simpson's paradox and explain the relevance of this paradox to the cross-tabulation of variables.

As explained in Chapter 1, data can be classified as either qualitative or quantitative. **Qualitative data** use labels or names to identify categories of like items. **Quantitative data** are numerical values that indicate how much or how many.

This chapter introduces tabular and graphical methods commonly used to summarize both qualitative and quantitative data. Everyone is exposed to these types of presentation in annual reports (see Statistics in Practice), newspaper articles and research studies. It is important to understand how they are prepared and how they should be interpreted. We begin with methods for summarizing single variables. Section 2.3 introduces methods for summarizing the relationship between two variables.

Modern spreadsheet and statistical software packages provide extensive capabilities for summarizing data and preparing graphical presentations. EXCEL, IBM SPSS and MINITAB are three widely available packages. There are guides to some of their capabilities on the companion website.

19

20 CHAPTER 2 DESCRIPTIVE STATISTICS: TABULAR AND GRAPHICAL PRESENTATIONS

STATISTICS IN PRACTICE
Marks & Spencer: not just any statistical graphics

Marks & Spencer has a company history going back to 1884. The group is based in London, but has offices across the UK as well as overseas. Most people are likely to have come across its promotional activities and its advertising slogan 'Your M&S'. Marks & Spencer advertisements have featured a long list of well-known faces, including Twiggy, Erin O'Connor, David Beckham, Claudia Schiffer, Rosie Huntington-Whiteley and Antonio Banderas.

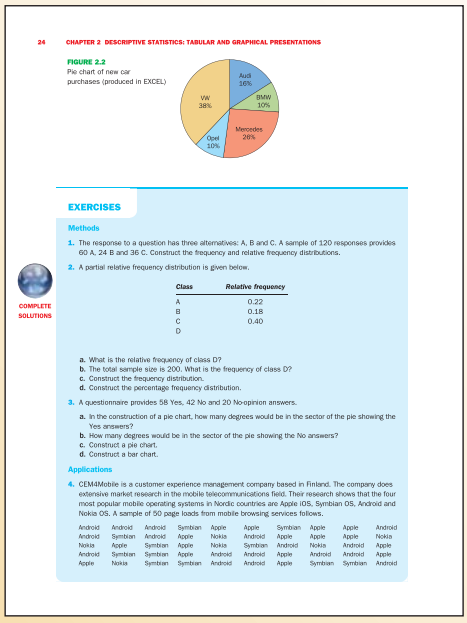
Marks & Spencer's shares are traded on the London Stock Exchange and it is a constituent of the FTSE 100 Index. Like all public companies, Marks & Spencer publishes an annual report. In the annual report, alongside many photographs of its ambassadors and models, there are pictures of a different nature: statistical charts illustrating in particular the financial performance of the company. The examples here are from Marks and Spencer's 2013 Annual Report. First is a chart showing Marks & Spencer's governance framework, then a bar chart showing the breakdown of Marks & Spencer's international revenue, and finally a line graph showing mystery shopper feedback.

We are exposed to statistical charts of this type almost daily: in newspapers and magazines, on TV, online and in business reports such as the Marks & Spencer Annual Report. In this chapter, you will learn about tabular and graphical methods of descriptive statistics such as frequency distributions, bar charts, histograms, stem-and-leaf displays, cross-tabulations and others. The goal of these methods is to summarize data so that they can be easily understood and interpreted.

A window display showing an array of personalities who have modelled for Marks & Spencer

Learning Objectives We have set out clear learning objectives at the start of each chapter in the text, as is now common in texts in the UK and elsewhere. These objectives summarize the core content of each chapter in a list of key points.

Statistics in Practice Each chapter begins with a Statistics in Practice article that describes an application of the statistical methodology to be covered in the chapter.



24 CHAPTER 2 DESCRIPTIVE STATISTICS: TABULAR AND GRAPHICAL PRESENTATIONS

FIGURE 2.2
Pie chart of new car purchases (produced in EXCEL)

EXERCISES

Methods

1. The response to a question has three alternatives: A, B and C. A sample of 120 responses provides 60 A, 24 B and 36 C. Construct the frequency and relative frequency distributions.

2. A partial relative frequency distribution is given below.

Class	*Relative frequency*
A	0.22
B	0.18
C	0.40
D	

COMPLETE SOLUTIONS

a. What is the relative frequency of class D?
b. The total sample size is 200. What is the frequency of class D?
c. Construct the frequency distribution.
d. Construct the percentage frequency distribution.

3. A questionnaire provides 58 Yes, 42 No and 20 No-opinion answers.

a. In the construction of a pie chart, how many degrees would be in the sector of the pie showing the Yes answers?
b. How many degrees would be in the sector of the pie showing the No answers?
c. Construct a pie chart.
d. Construct a bar chart.

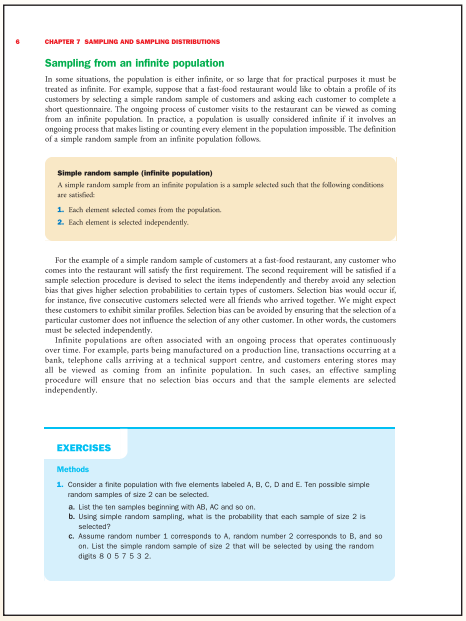
Applications

4. CEM4Mobile is a customer experience management company based in Finland. The company does extensive market research in the mobile telecommunications field. Their research shows that the four most popular mobile operating systems in Nordic countries are Apple IOS, Symbian OS, Android and Nokia OS. A sample of 50 page loads from mobile browsing services follows.

Android	Android	Android	Symbian	Apple	Apple	Symbian	Apple	Apple	Apple	Android
Android	Symbian	Android	Apple	Nokia	Android	Apple	Apple	Apple	Nokia	
Nokia	Apple	Symbian	Apple	Nokia	Symbian	Android	Nokia	Android	Apple	
Android	Symbian	Symbian	Apple	Android	Android	Apple	Android	Android	Apple	
Apple	Nokia	Symbian	Symbian	Android	Android	Apple	Symbian	Symbian	Android	

Exercises The exercises are split into two parts: Methods and Applications. The Methods exercises require students to use the formulae and make the necessary computations. The Applications exercises require students to use the chapter material in real-world situations. Thus, students first focus on the computational 'nuts and bolts', then move on to the subtleties of statistical application and interpretation. Answers to even-numbered exercises are provided on the online platform, while a full set of answers are provided in the lecturers' Solutions Manual. Supplementary exercises are provided on the textbook's online platform. Self-test exercises are highlighted throughout by the 'COMPLETE SOLUTIONS' icon and contain fully-worked solutions on the online platform.

COMPLETE SOLUTIONS



6 CHAPTER 7 SAMPLING AND SAMPLING DISTRIBUTIONS

Sampling from an infinite population

In some situations, the population is either infinite, or so large that for practical purposes it must be treated as infinite. For example, suppose that a fast-food restaurant would like to obtain a profile of its customers by selecting a simple random sample of customers and asking each customer to complete a short questionnaire. The ongoing process of customer visits to the restaurant can be viewed as coming from an infinite population. In practice, a population is usually considered infinite if it involves an ongoing process that makes listing or counting every element in the population impossible. The definition of a simple random sample from an infinite population follows.

Simple random sample (infinite population)

A simple random sample from an infinite population is a sample selected such that the following conditions are satisfied:

1. Each element selected comes from the population.
2. Each element is selected independently.

For the example of a simple random sample of customers at a fast-food restaurant, any customer who comes into the restaurant will satisfy the first requirement. The second requirement will be satisfied if a sample selection procedure is devised to select the items independently and thereby avoid any selection bias that gives higher selection probabilities to certain types of customers. Selection bias would occur if, for instance, five consecutive customers selected were all friends who arrived together. We might expect these customers to exhibit similar profiles. Selection bias can be avoided by ensuring that the selection of a particular customer does not influence the selection of any other customer. In other words, the customers must be selected independently.

Infinite populations are often associated with an ongoing process that operates continuously over time. For example, parts being manufactured on a production line, transactions occurring at a bank, telephone calls arriving at a technical support centre, and customers entering stores may all be viewed as coming from an infinite population. In such cases, an effective sampling procedure will ensure that no selection bias occurs and that the sample elements are selected independently.

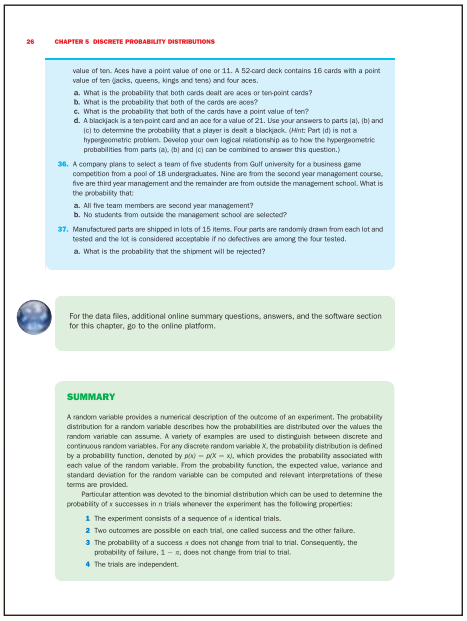
EXERCISES

Methods

1. Consider a finite population with five elements labeled A, B, C, D and E. Ten possible simple random samples of size 2 can be selected.
   a. List the ten samples beginning with AB, AC and so on.
   b. Using simple random sampling, what is the probability that each sample of size 2 is selected?
   c. Assume random number 1 corresponds to A, random number 2 corresponds to B, and so on. List the simple random sample of size 2 that will be selected by using the random digits 8 0 5 7 5 3 2.

Notes Recent US editions have included marginal and end-of-chapter notes.

We have not adopted this layout, but have included the important material in the text itself.



26 CHAPTER 5 DISCRETE PROBABILITY DISTRIBUTIONS

value of ten. Aces have a point value of one or 11. A 52-card deck contains 16 cards with a point value of ten (jacks, queens, kings and tens) and four aces.

a. What is the probability that both cards dealt are aces or ten-point cards?
b. What is the probability that both of the cards are aces?
c. What is the probability that both of the cards have a point value of ten?
d. A blackjack is a ten-point card and an ace for a value of 21. Use your answers to parts (a), (b) and (c) to determine the probability that a player is dealt a blackjack. (Hint: Part (d) is not a hypergeometric problem. Develop your own logical relationship as to how the hypergeometric probabilities from parts (a), (b) and (c) can be combined to answer this question.)

36. A company plans to select a team of five students from Gulf university for a business game competition from a pool of 18 undergraduates. Nine are from the second year management course, five are third year management and the remainder are from outside the management school. What is the probability that:
    a. All five team members are second year management?
    b. No students from outside the management school are selected?

37. Manufactured parts are shipped in lots of 15 items. Four parts are randomly drawn from each lot and tested and the lot is considered acceptable if no defectives are among the four tested.
    a. What is the probability that the shipment will be rejected?

For the data files, additional online summary questions, answers, and the software section for this chapter, go to the online platform.

SUMMARY

A random variable provides a numerical description of the outcome of an experiment. The probability distribution for a random variable describes how the probabilities are distributed over the values the random variable can assume. A variety of examples are used to distinguish between discrete and continuous random variables. For any discrete random variable X, the probability distribution is defined by a probability function, denoted by $p(x) = p(X = x)$, which provides the probability associated with each value of the random variable. From the probability function, the expected value, variance and standard deviation for the random variable can be computed and relevant interpretations of these terms are provided.

Particular attention was devoted to the binomial distribution which can be used to determine the probability of x successes in n trials whenever the experiment has the following properties:

1 The experiment consists of a sequence of n identical trials.
2 Two outcomes are possible on each trial, one called success and the other failure.
3 The probability of a success π does not change from trial to trial. Consequently, the probability of failure, $1 - \pi$, does not change from trial to trial.
4 The trials are independent.

Summaries Each chapter includes a summary to remind students of what they have learnt so far and offer a useful way to review for exams.

18 CHAPTER 1 DATA AND STATISTICS

For the data files and additional online resources for Chapter 1, go to the accompanying online platform. (See the 'About the Digital Resources' page in the front of the book for more information on access.)

SUMMARY

Statistics is the art and science of collecting, analyzing, presenting and interpreting data. Nearly every college student majoring in business or economics is required to take a course in statistics. We began the chapter by describing typical statistical applications for business and economics.

Data consist of the facts and figures that are collected and analyzed. A set of measurements obtained for a particular element is an observation. Four scales of measurement used to obtain data on a particular variable include nominal, ordinal, interval and ratio. The scale of measurement for a variable is nominal when the data use labels or names to identify an attribute of an element. The scale is ordinal if the data demonstrate the properties of nominal data and the order or rank of the data is meaningful. The scale is interval if the data demonstrate the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Finally, the scale of measurement is ratio if the data show all the properties of interval data and the ratio of two values is meaningful.

For purposes of statistical analysis, data can be classified as categorical or quantitative.

Categorical data use labels or names to identify an attribute of each element. Categorical data use either the nominal or ordinal scale of measurement and may be non-numeric or numeric. Quantitative data are numeric values that indicate how much or how many. Quantitative data use either the interval or ratio scale of measurement. Ordinary arithmetic operations are meaningful only if the data are quantitative. Therefore, statistical computations used for quantitative data are not always appropriate for categorical data.

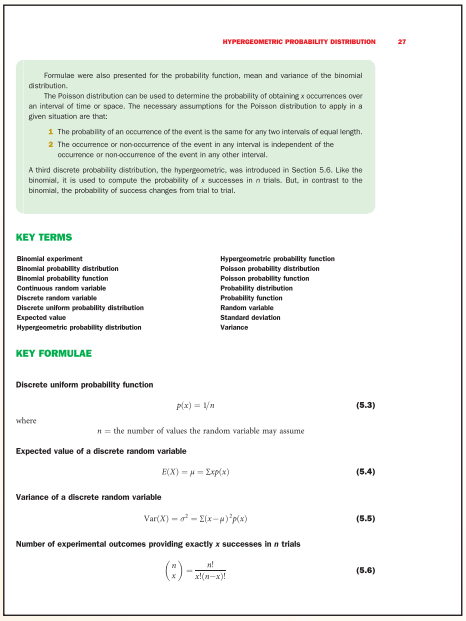
In Sections 1.4 and 1.5 we introduced the topics of descriptive statistics and statistical inference. Definitions of the population and sample were provided and different types of descriptive statistics – tabular, graphical, and numerical – used to summarize data. The process of statistical inference uses data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.

The last two sections of the chapter provide information on the role of computers in statistical analysis and a brief overview of the relative new field of data mining.

KEY TERMS

Categorical data
Categorical variable
Census
Cross-sectional data
Data
Data mining
Data set
Descriptive statistics
Elements
Interval scale
Nominal scale
Observation
Ordinal scale
Population
Quantitative data
Quantitative variable
Ratio scale
Sample
Sample survey
Statistical inference
Statistics
Time series data
Variable

Data sets accompany text Over 200 data sets are available on the online platform that accompanies the text. The data sets are available in MINITAB, SPSS and EXCEL formats. Data set logos are used in the text to identify the data sets that are available online. Data sets for all case problems as well as data sets for larger exercises are also included on the online platform.



HYPERGEOMETRIC PROBABILITY DISTRIBUTION 27

Formulae were also presented for the probability function, mean and variance of the binomial distribution.

The Poisson distribution can be used to determine the probability of obtaining x occurrences over an interval of time or space. The necessary assumptions for the Poisson distribution to apply in a given situation are that:

1. The probability of an occurrence of the event is the same for any two intervals of equal length.
2. The occurrence or non-occurrence of the event in any interval is independent of the occurrence or non-occurrence of the event in any other interval.

A third discrete probability distribution, the hypergeometric, was introduced in Section 5.6. Like the binomial, it is used to compute the probability of x successes in n trials. But, in contrast to the binomial, the probability of success changes from trial to trial.

KEY TERMS

Binomial experiment
Binomial probability distribution
Binomial probability function
Continuous random variable
Discrete random variable
Discrete uniform probability distribution
Expected value
Hypergeometric probability distribution
Hypergeometric probability function
Poisson probability distribution
Poisson probability function
Probability distribution
Probability function
Random variable
Standard deviation
Variance

KEY FORMULAE

Discrete uniform probability function

$$p(x) = 1/n \tag{5.3}$$

where

n = the number of values the random variable may assume

Expected value of a discrete random variable

$$E(X) = \mu = \Sigma x p(x) \tag{5.4}$$

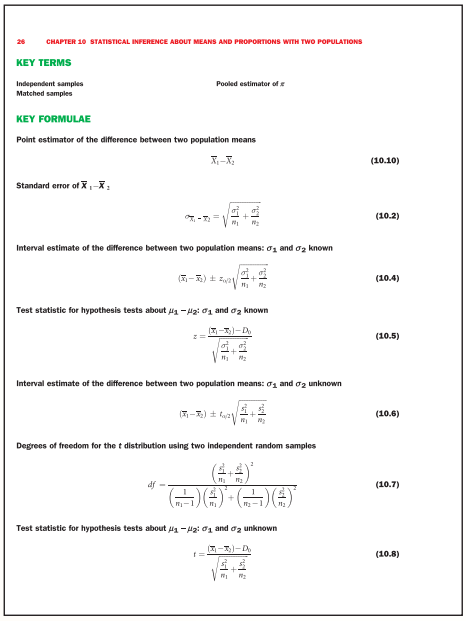
Variance of a discrete random variable

$$\text{Var}(X) = \sigma^2 = \Sigma (x-\mu)^2 p(x) \tag{5.5}$$

Number of experimental outcomes providing exactly x successes in n trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

Key terms Key terms are highlighted in the text, listed at the end of each chapter and given a full definition in the Glossary at the end of the textbook.



26 CHAPTER 10 STATISTICAL INFERENCE ABOUT MEANS AND PROPORTIONS WITH TWO POPULATIONS

KEY TERMS

Independent samples
Matched samples
Pooled estimator of π

KEY FORMULAE

Point estimator of the difference between two population means

$$\bar{X}_1 - \bar{X}_2 \tag{10.10}$$

Standard error of $\bar{X}_1 - \bar{X}_2$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{10.2}$$

Interval estimate of the difference between two population means: σ_1 and σ_2 known

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{10.4}$$

Test statistic for hypothesis tests about $\mu_1 - \mu_2$: σ_1 and σ_2 known

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{10.5}$$

Interval estimate of the difference between two population means: σ_1 and σ_2 unknown

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{10.6}$$

Degrees of freedom for the t distribution using two independent random samples

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1 - 1}\right)\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2 - 1}\right)\left(\frac{s_2^2}{n_2}\right)^2} \tag{10.7}$$

Test statistic for hypothesis tests about $\mu_1 - \mu_2$: σ_1 and σ_2 unknown

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{10.8}$$

Key formulae Key formulae are listed at the end of each chapter for easy reference.

THE WEIGHTED MEAN AND WORKING WITH GROUPED DATA 85

CASE PROBLEM 2

Chocolate Perfection Website Transactions

Chocolate Perfection manufactures and sells quality chocolate products in Dubai. Two years ago the company developed a website and began selling its products over the Internet. Website sales have exceeded the company's expectations, and management is now considering strategies to increase sales even further. To learn more about the website customers, a sample of 50 Chocolate Perfection transactions was selected from the previous month's sales. Data showing the day of the week each transaction was made, the type of browser the customer used, the time spent on the website, the number of website pages viewed and the amount spent by each of the 50 customers are contained in the file named 'Shoppers'. Amount spent is in United Arab Emirates dirham (AED). (One Euro is around five AED.) A screenshot of a data is shown below.

Customer	Day	Browser	Time (min)	Pages Viewed	Amount Spent (AED)
1	Mon	Internet Explorer	12.0	4	200.09
2	Wed	Other	19.5	6	348.28
3	Mon	Internet Explorer	8.5	4	97.92
4	Tue	Firefox	11.4	2	164.16
5	Wed	Internet Explorer	11.3	4	243.21
6	Sat	Firefox	10.5	6	248.83
7	Sun	Internet Explorer	11.4	2	132.27
8	Fri	Firefox	4.3	6	205.37
9	Wed	Firefox	12.7	3	260.35

SHOPPERS

Chocolate Perfection would like to use the sample data to determine if online shoppers who spend more time and view more pages also spend more money during their visit to the website. The company would also like to investigate the effect that the day of the week and the type of browser has on sales.

Managerial report

Use the methods of descriptive statistics to learn about the customers who visit the Chocolate Perfection website. Include the following in your report:

1. Graphical and numerical summaries for the length of time the shopper spends on the website, the number of pages viewed and the mean amount spent per transaction. Discuss what you learn about Chocolate Perfection's online shoppers from these numerical summaries.
2. Summarize the frequency, the total amount spent and the mean amount spent per transaction for each day of week. What observations can you make about Chocolate Perfection's business based on the day of the week? Discuss.
3. Summarize the frequency, the total amount spent and the mean amount spent per transaction for each type of browser. What observations can you make about Chocolate Perfection's business, based on the type of browser? Discuss.
4. Construct a scatter diagram and compute the sample correlation coefficient to explore the relationship between the time spent on the website and the amount spent. Use the horizontal axis for the time spent on the website. Discuss.
5. Construct a scatter diagram and compute the sample correlation coefficient to explore the relationship between the number of website pages viewed and the amount spent. Use the horizontal axis for the number of website pages viewed. Discuss.
6. Construct a scatter diagram and compute the sample correlation coefficient to explore the relationship between the time spent on the website and the number of pages viewed. Use the horizontal axis to represent the number of pages viewed. Discuss.

Case problems The end-of-chapter case problems provide students with the opportunity to analyse somewhat larger data sets and prepare managerial reports based on the results of the analysis.

DIGITAL RESOURCES

Dedicated Instructor Resources

To discover the dedicated instructor online support resources accompanying this textbook, instructors should register here for access: **http://login.cengage.com**

Resources include:

- Solutions Manual
- ExamView Testbank
- PowerPoint slides

Instructor access

Instructors can access the online student platform by registering at **http://login.cengage.com** or by speaking to their local Cengage Learning EMEA representative.

Instructor resources

Instructors can use the integrated Engagement Tracker to track students' preparation and engagement. The tracking tool can be used to monitor progress of the class as a whole, or for individual students.

Student access

Students can access the online platform using the unique personal access card included in the front of the book.

Student resources

The platform offers a range of interactive learning tools tailored to the third edition of *Statistics for Business and Economics*, including:

- Interactive eBook
- Data files referred to in the text
- Answers to in-text exercises
- Software section
- Four additional chapters for further study
- Glossary, flashcards and more

1 Data and Statistics

CHAPTER CONTENTS

LEARNING OBJECTIVES

After reading this chapter and doing the exercises, you should be able to:

1 Appreciate the breadth of statistical applications in business and economics.

2 Understand the meaning of the terms elements, variables and observations, as they are used in statistics.

3 Understand the difference between qualitative, quantitative, cross-sectional and time series data.

4 Find out about data sources available for statistical analysis both internal and external to the firm.

5 Appreciate how errors can arise in data.

6 Understand the meaning of descriptive statistics and statistical inference.

7 Distinguish between a population and a sample.

8 Understand the role a sample plays in making statistical inferences about the population.

Frequently, we see the following kinds of statements in newspaper and magazine articles:

- The Ifo World Economic Climate Index fell again substantially in January 2009. The climate indicator stands at 50.1 (1995 = 100); its historically lowest level since introduction in the early 1980s (CESifo, April 2009).
- The IMF projected the global economy would shrink 1.3 per cent in 2009 (*Fin24*, 23 April 2009).
- The Footsie finished the week on a winning streak despite shock figures that showed the economy has contracted by almost 2 per cent already in 2009 (*This is Money*, 25 April 2009).
- China's growth rate fell to 6.1 per cent in the year to the first quarter (*The Economist*, 16 April 2009).

- GM receives further $2bn in loans (*BBC News*, 24 April 2009).
- Handset shipments to drop by 20 per cent (*In-Stat*, 2009).

The numerical facts in the preceding statements (50.1, 1.3 per cent, 2 per cent, 6.1 per cent, $2bn, 20 per cent) are called statistics. Thus, in everyday usage, the term *statistics* refers to numerical facts. However, the field, or subject, of statistics involves much more than numerical facts. In a broad sense, **statistics** is the art and science of collecting, analyzing, presenting and interpreting data. Particularly in business and economics, the information provided by collecting, analyzing, presenting and interpreting data gives managers and decision-makers a better understanding of the business and economic environment and thus enables them to make more informed and better decisions. In this text, we emphasize the use of statistics for business and economic decision-making.

Chapter 1 begins with some illustrations of the applications of statistics in business and economics. In Section 1.2 we define the term *data* and introduce the concept of a data set. This section also introduces key terms such as *variables* and *observations*, discusses the difference between quantitative and categorical data, and illustrates the uses of cross-sectional and time series data. Section 1.3 discusses how data can be obtained from existing sources or through survey and experimental studies designed to obtain new data. The important role that the Internet now plays in obtaining data is also highlighted. The use of data in developing descriptive statistics and in making statistical inferences is described in Sections 1.4 and 1.5. The last two sections of Chapter 1 outline respectively the role of computers in statistical analysis and introduce the relatively new field of data mining.

STATISTICS IN PRACTICE
The Economist

Founded in 1843, *The Economist* is an international weekly news and business magazine written for top-level business executives and political decision-makers. The publication aims to provide readers with in-depth analyses of international politics, business news and trends, global economics and culture.

The Economist is published by the Economist Group – an international company employing nearly 1000 staff worldwide – with offices in London, Frankfurt, Paris and Vienna; in New York, Boston and Washington, DC; and in Hong Kong, mainland China, Singapore and Tokyo.

Between 1998 and 2008 the magazine's worldwide circulation grew by 100 per cent – recently exceeding 180 000 in the UK, 230 000 in continental Europe, 780 000 plus copies in North America and nearly 130 000 in the Asia-Pacific region. It is read in more than 200 countries and with a readership of four million, is one of the world's most influential business publications. Along with the *Financial Times*, it is arguably one of the two most successful print publications to be introduced in the US market during the past decade.

Complementing *The Economist* brand within the Economist Brand family, the Economist Intelligence Unit provides access to a comprehensive database of worldwide indicators and forecasts covering more than 200 countries, 45 regions and eight key industries. The Economist Intelligence Unit aims to help executives make informed business decisions through dependable intelligence delivered online, in print, in customized research as well as through conferences and peer interchange.

Alongside the Economist Brand family, the Group manages and runs the CFO and Government brand families for the benefit of senior finance executives and government decision-makers (in Brussels and Washington respectively).

1.1 APPLICATIONS IN BUSINESS AND ECONOMICS

In today's global business and economic environment, anyone can access vast amounts of statistical information. The most successful managers and decision-makers understand the information and know how to use it effectively. In this section, we provide examples that illustrate some of the uses of statistics in business and economics.

Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients. For instance, suppose an accounting firm wants to determine whether the amount of accounts receivable shown on a client's balance sheet fairly represents the actual amount of accounts receivable. Usually the large number of individual accounts receivable makes reviewing and validating every account too time-consuming and expensive. As common practice in such situations, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

Finance

Financial analysts use a variety of statistical information to guide their investment recommendations. In the case of stocks, the analysts review a variety of financial data including price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, a financial analyst can begin to draw a conclusion as to whether an individual stock is over- or under-priced. Similarly, historical trends in stock prices can provide a helpful indication on when investors might consider entering (or re-entering) the market. For example, *Money Week* (3 April 2009) reported a Goldman Sachs analysis that indicated, because stocks were unusually cheap at the time, real average returns of up to 6 per cent in the US and 7 per cent in Britain might be possible over the next decade – based on long-term cyclically adjusted price/earnings ratios.

Marketing

Electronic scanners at retail checkout counters collect data for a variety of marketing research applications. For example, data suppliers such as ACNielsen purchase point-of-sale scanner data from grocery stores, process the data and then sell statistical summaries of the data to manufacturers. Manufacturers spend vast amounts per product category to obtain this type of scanner data. Manufacturers also purchase data and statistical summaries on promotional activities such as special pricing and the use of in-store displays. Brand managers can review the scanner statistics and the promotional activity statistics to gain a better understanding of the relationship between promotional activities and sales. Such analyses often prove helpful in establishing future marketing strategies for the various products.

Production

Today's emphasis on quality makes quality control an important application of statistics in production. A variety of statistical quality control charts are used to monitor the output of a production process. In particular, an x-bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 330g of a soft drink. Periodically, a production worker selects a sample of containers and computes the average number of grams in the sample. This average, or x-bar value, is plotted on an x-bar chart. A plotted value above the chart's upper control limit indicates overfilling, and a plotted value below the chart's lower control limit indicates underfilling. The process is termed 'in control' and allowed to continue as long as the plotted x-bar values fall between the chart's upper and lower control limits. Properly interpreted, an x-bar chart can help determine when adjustments are necessary to correct a production process.

Economics

Economists frequently provide forecasts about the future of the economy or some aspect of it. They use a variety of statistical information in making such forecasts. For instance, in forecasting inflation rates, economists use statistical information on such indicators as the Producer Price Index, the unemployment rate and manufacturing capacity utilization. Often these statistical indicators are entered into computerized forecasting models that predict inflation rates.

Applications of statistics such as those described in this section are an integral part of this text. Such examples provide an overview of the breadth of statistical applications. To supplement these examples, chapter-opening Statistics in Practice articles obtained from a variety of topical sources are used to introduce the material covered in each chapter. These articles show the importance of statistics in a wide variety of business and economic situations.

1.2 DATA

Data are the facts and figures collected, analyzed and summarized for presentation and interpretation. All the data collected in a particular study are referred to as the **data set** for the study. Table 1.1 shows a data set summarizing information for equity (share) trading at the 22 European Stock Exchanges in March 2009.

EXCHANGES 2009

TABLE 1.1 European stock exchange monthly statistics domestic equity trading (electronic order book transactions) March 2009

	Total	
Exchange	**Trades**	**Turnover**
Athens	599 192	2 009.8
Borsa Italiana	5 921 099	44 385.9
Bratislava	111	0.1
Bucharest	79 921	45.3
Budapest	298 871	1 089.6
Bulgarian	14 040	64.4
Cyprus	31 167	76.1
Deutsche Börse	7 642 241	86 994.5
Euronext	15 282 996	116 488
Irish	79 973	549.8
Ljublijana	11 172	35.6
London	16 539 588	114 283.6
Luxembourg	1 152	125
Malta	638	1.9
NASDAQ OMX Nordic	4 550 073	40 927.4
Oslo Bars	981 362	9 755.1
Prague	65 153	1 034.8
SIX Swiss	440 578	2 667.1
Spanish (BME)	2 799 329	60 387.6
SWX Europe	n/a	n/a
Warsaw	1 155 379	2 468.6
Wiener Borse	433 545	2 744
TOTAL	**56 927 580**	**486 021.7**

Source: European Stock Exchange monthly statistics (www.fese.be/en/?inc=art&id=3)

Elements, variables and observations

Elements are the entities on which data are collected. For the data set in Table 1.1, each individual European exchange is an element; the element names appear in the first column. With 22 exchanges, the data set contains 22 elements.

A **variable** is a characteristic of interest for the elements. The data set in Table 1.1 includes the following three variables:

- *Exchange*: at which the equities were traded.
- *Trades*: number of trades during the month.
- *Turnover*: value of trades (€m) during the month.

Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an **observation**. Referring to Table 1.1, we see that the set of measurements for the first observation (Athens Exchange) is 599 192 and 2009.8. The set of measurements for the second observation (Borsa Italiana) is 5 921 099 and 44 385.9; and so on. A data set with 22 elements contains 22 observations.

Scales of measurement

Data collection requires one of the following scales of measurement: nominal, ordinal, interval or ratio. The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

When the data for a variable consist of labels or names used to identify an attribute of the element, the scale of measurement is considered a **nominal scale**. For example, referring to the data in Table 1.1, we see that the scale of measurement for the exchange variable is nominal because Athens Exchange, Borsa Italiana … Wiener Börse are labels used to identify where the equities are traded. In cases where the scale of measurement is nominal, a numeric code as well as non-numeric labels may be used. For example, to facilitate data collection and to prepare the data for entry into a computer database, we might use a numeric code by letting 1, denote the Athens Exchange, 2, the Borsa Italiana … and 22, Wiener Börse. In this case the numeric values 1, 2, … 22 provide the labels used to identify where the stock is traded. The scale of measurement is nominal even though the data appear as numeric values.

The scale of measurement for a variable is called an **ordinal scale** if the data exhibit the properties of nominal data and the order or rank of the data is meaningful. For example, Eastside Automotive sends customers a questionnaire designed to obtain data on the quality of its automotive repair service. Each customer provides a repair service rating of excellent, good or poor. Because the data obtained are the labels – excellent, good or poor – the data have the properties of nominal data. In addition, the data can be ranked, or ordered, with respect to the service quality. Data recorded as excellent indicate the best service, followed by good and then poor. Thus, the scale of measurement is ordinal. Note that the ordinal data can also be recorded using a numeric code. For example, we could use 1 for excellent, 2 for good and 3 for poor to maintain the properties of ordinal data. Thus, data for an ordinal scale may be either non-numeric or numeric.

The scale of measurement for a variable becomes an **interval scale** if the data show the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numeric. Graduate Management Admission Test (GMAT) scores are an example of interval-scaled data. For example, three students with GMAT scores of 620 550 and 470 can be ranked or ordered in terms of best performance to poorest performance. In addition, the differences between the scores are meaningful. For instance, student one scored 620 – 550 = 70 points more than student two, while student two scored 550 – 470 = 80 points more than student three.

The scale of measurement for a variable is a **ratio scale** if the data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point. For example, consider the cost of a car. A zero value for the cost would

indicate that the car has no cost and is free. In addition, if we compare the cost of €30 000 for one car to the cost of €15 000 for a second car, the ratio property shows that the first car is €30 000/€15 000 = two times, or twice, the cost of the second car.

Categorical and quantitative data

Data can be further classified as either categorical or quantitative. **Categorical data** include labels or names used to identify an attribute of each element. Categorical data use either the nominal or ordinal scale of measurement and may be non-numeric or numeric. **Quantitative data** require numeric values that indicate how much or how many. Quantitative data are obtained using either the interval or ratio scale of measurement.

A **categorical variable** is a variable with categorical data, and a **quantitative variable** is a variable with quantitative data. The statistical analysis appropriate for a particular variable depends upon whether the variable is categorical or quantitative. If the variable is categorical, the statistical analysis is rather limited. We can summarize categorical data by counting the number of observations in each category or by computing the proportion of the observations in each category. However, even when the categorical data use a numeric code, arithmetic operations such as addition, subtraction, multiplication and division do not provide meaningful results. Section 2.1 discusses ways for summarizing categorical data.

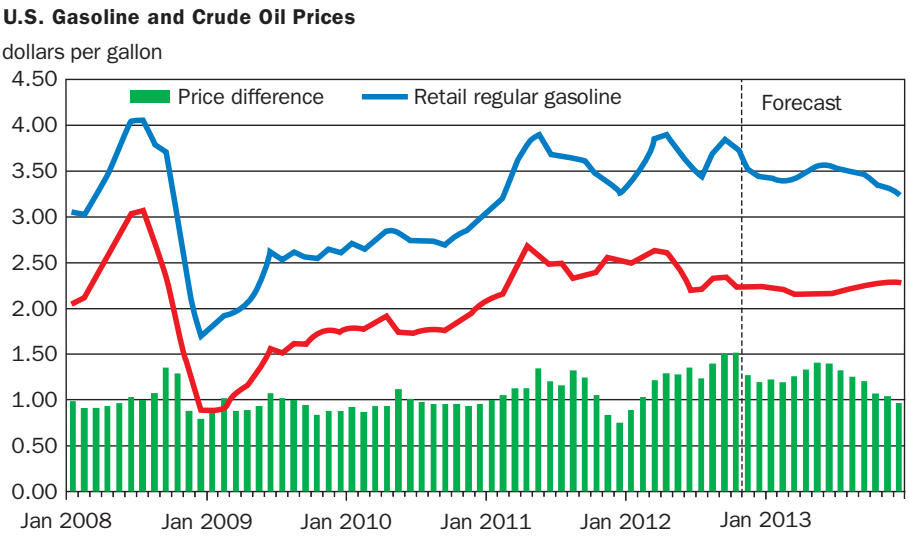
On the other hand, arithmetic operations often provide meaningful results for a quantitative variable. For example, for a quantitative variable, the data may be added and then divided by the number of observations to compute the average value. This average is usually meaningful and easily interpreted. In general, more alternatives for statistical analysis are possible when the data are quantitative. Section 2.2 and Chapter 3 provide ways of summarizing quantitative data.

Cross-sectional and time series data

For purposes of statistical analysis, distinguishing between cross-sectional data and time series data is important. **Cross-sectional data** are data collected at the same or approximately the same point in time. The data in Table 1.1 are cross-sectional because they describe the two variables for the 22 exchanges at the same point in time. **Time series data** are data collected over several time periods. For example, Figure 1.1 provides a graph of the wholesale price (US$) of crude oil per gallon for the period January 2008 and January 2012. It shows that starting around July 2008 the average price dipped sharply to less than $2 per gallon. However, by November 2011 it had recovered to $3 per gallon since when it has mostly hovered between $3.50 and $4 per gallon. Most of the statistical methods presented in this text apply to cross-sectional rather than time series data.

Quantitative data that measure how many are discrete. Quantitative data that measure how much are continuous because no separation occurs between the possible data values.

FIGURE 1.1
Wholesale price of crude oil per gallon (US$) 2008–2012
EIA (www.eia.doe.gov/)



Crude oil price is composite refiner acquisition cost. Retail prices include state and federal

Source: Short-Term Energy Outlook, November 2012

1.3 DATA SOURCES

Data can be obtained from existing sources or from surveys and experimental studies designed to collect new data.

Existing sources

In some cases, data needed for a particular application already exist. Companies maintain a variety of databases about their employees, customers and business operations. Data on employee salaries, ages and years of experience can usually be obtained from internal personnel records. Other internal records contain data on sales, advertising expenditures, distribution costs, inventory levels and production quantities. Most companies also maintain detailed data about their customers. Table 1.2 shows some of the data commonly available from internal company records.

Organizations that specialize in collecting and maintaining data make available substantial amounts of business and economic data. Companies access these external data sources through leasing arrangements or by purchase. Dun & Bradstreet, Bloomberg and the Economist Intelligence Unit are three sources that provide extensive business database services to clients. ACNielsen built successful businesses collecting and processing data that they sell to advertisers and product manufacturers.

Data are also available from a variety of industry associations and special interest organizations. The European Tour Operators, Association and European Travel Commission provide information on tourist trends and travel expenditures by visitors to and from countries in Europe. Such data would be of interest to firms and individuals in the travel industry. The Graduate Management Admission Council maintains data on test scores, student characteristics and graduate management education programmes. Most of the data from these types of sources are available to qualified users at a modest cost.

The Internet continues to grow as an important source of data and statistical information. Almost all companies maintain websites that provide general information about the company as well as data on sales, number of employees, number of products, product prices and product specifications. In addition, a number of companies now specialize in making information available over the Internet. As a result, one can obtain access to stock quotes, meal prices at restaurants, salary data and an almost infinite variety of information. Government agencies are another important source of existing data. For instance, Eurostat maintains considerable data on employment rates, wage rates, size of the labour force and union membership. Table 1.3 lists selected governmental agencies and some of the data they provide. Most government agencies that collect and process data also make the results available through a website. For instance, the Eurostat has a wealth of data at its website, http://ec.europa.eu/eurostat. Figure 1.2 shows the homepage for the Eurostat.

TABLE 1.2 Examples of data available from internal company records

Source	Some of the data typically available
Employee records	Name, address, social security number, salary, number of vacation days, number of sick days and bonus
Production records	Part or product number, quantity produced, direct labour cost and materials cost
Inventory records	Part or product number, number of units on hand, reorder level, economic order quantity and discount schedule
Sales records	Product number, sales volume, sales volume by region and sales volume by customer type
Credit records	Customer name, address, phone number, credit limit and accounts receivable balance
Customer profile	Age, gender, income level, household size, address and preferences

TABLE 1.3 Examples of data available from selected European sources

Source	Some of the data available
Europa rates (http://europa.eu)	Travel, VAT (value added tax), euro exchange
Eurostat (http://epp.eurostat.ec.europa.eu/)	employment, population and social conditions Education and training, labour market, living conditions and welfare
European Central Bank (www.ecb.int/)	Monetary, financial markets, interest rate and balance of payments statistics, unit labour costs, compensation per employee, labour productivity, consumer prices, construction prices

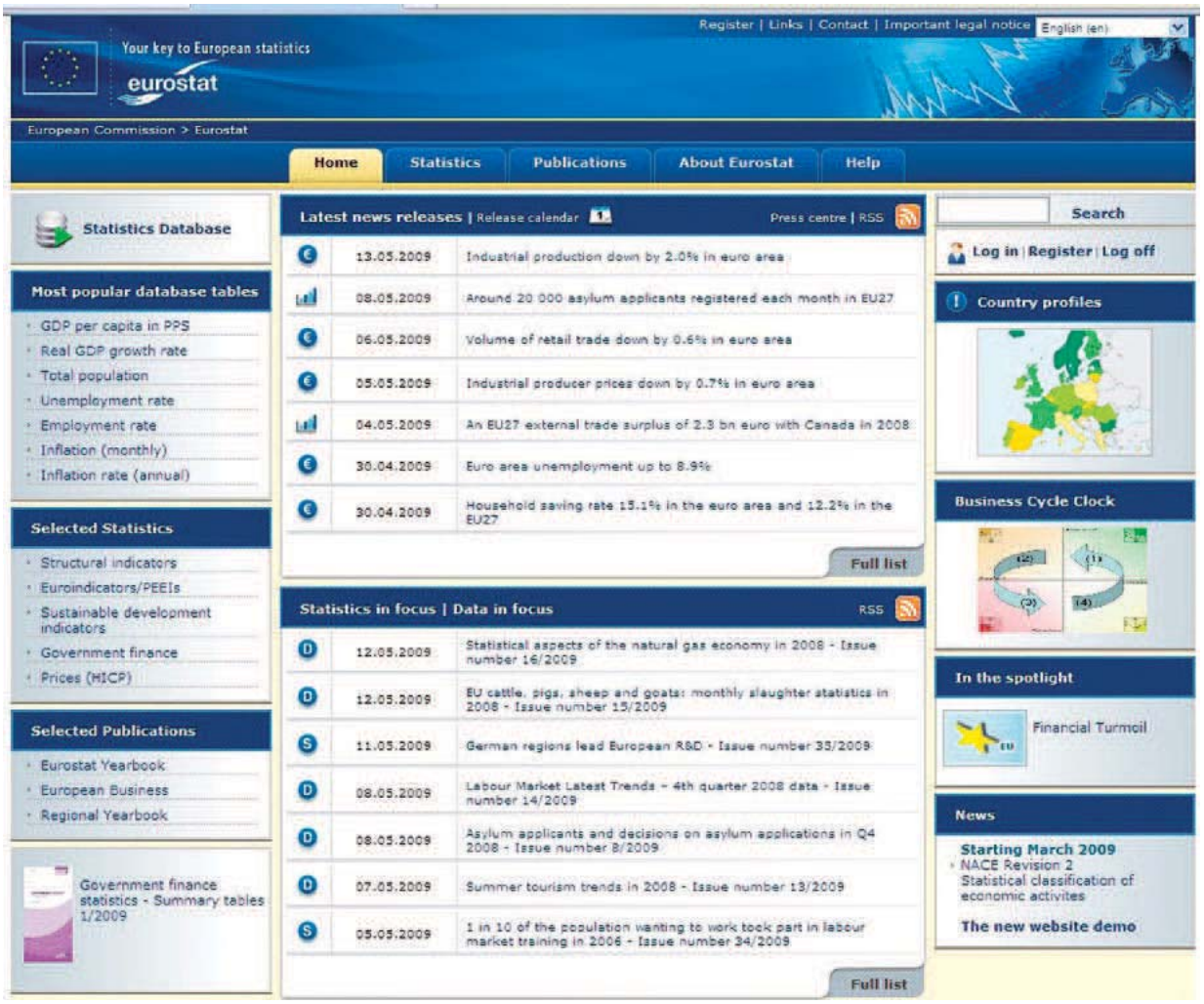


FIGURE 1.2
Eurostat homepage

Statistical studies

Sometimes the data needed for a particular application are not available through existing sources. In such cases, the data can often be obtained by conducting a statistical study. Statistical studies can be classified as either *experimental* or *observational.*

In an experimental study, a variable of interest is first identified. Then one or more other variables are identified and controlled so that data can be obtained about how they influence the variable of interest. For example, a pharmaceutical firm might be interested in conducting an experiment to learn about how a new drug affects blood pressure. Blood pressure is the variable of interest in the study. The dosage level of the new drug is another variable that is hoped to have a causal effect on blood pressure. To obtain data about the effect of the new drug, researchers select a sample of individuals. The dosage level of the new drug is controlled, as different groups of individuals are given different dosage levels. Before and after data on blood pressure are collected for each group. Statistical analysis of the experimental data can help determine how the new drug affects blood pressure.

Non-experimental, or observational, statistical studies make no attempt to control the variables of interest. A survey is perhaps the most common type of observational study. For instance, in a personal interview survey, research questions are first identified. Then a questionnaire is designed and administered to a sample of individuals. Some restaurants use observational studies to obtain data about their customers' opinions of the quality of food, service, atmosphere and so on. A questionnaire used by the Lobster Pot Restaurant in Limerick City, Ireland, is shown in Figure 1.3. Note that the customers completing the questionnaire are asked to provide ratings for five variables: food quality, friendliness of service, promptness of service, cleanliness and management. The response categories of excellent, good, satisfactory and unsatisfactory provide ordinal data that enable Lobster Pot's managers to assess the quality of the restaurant's operation.

Managers wanting to use data and statistical analyses as an aid to decision-making must be aware of the time and cost required to obtain the data. The use of existing data sources is desirable when data must be obtained in a relatively short period of time.

The LOBSTER Pot
RESTAURANT

We are happy you stopped by the Lobster Pot Restaurant and want to make sure you will come back. So, if you have a little time, we will really appreciate it if you will fill out this card. Your comments and suggestions are extremely important to us. Thank you!

Server's Name _____

	Excellent	Good	Satisfactory	Unsatisfactory
Food Quality	❑	❑	❑	❑
Friendly Service	❑	❑	❑	❑
Prompt Service	❑	❑	❑	❑
Cleanliness	❑	❑	❑	❑
Management	❑	❑	❑	❑

Comments _____

What prompted your vist to us? _____

Please drop in suggestion box at entrance. Thank you.

FIGURE 1.3
Customer opinion questionnaire used by the Lobster Pot Restaurant, Limerick City, Ireland