

Berenson O'Brien
Levine Jayne
Szabat Watson



Basic BUSINESS

Statistics

Concepts and Applications

4TH EDITION

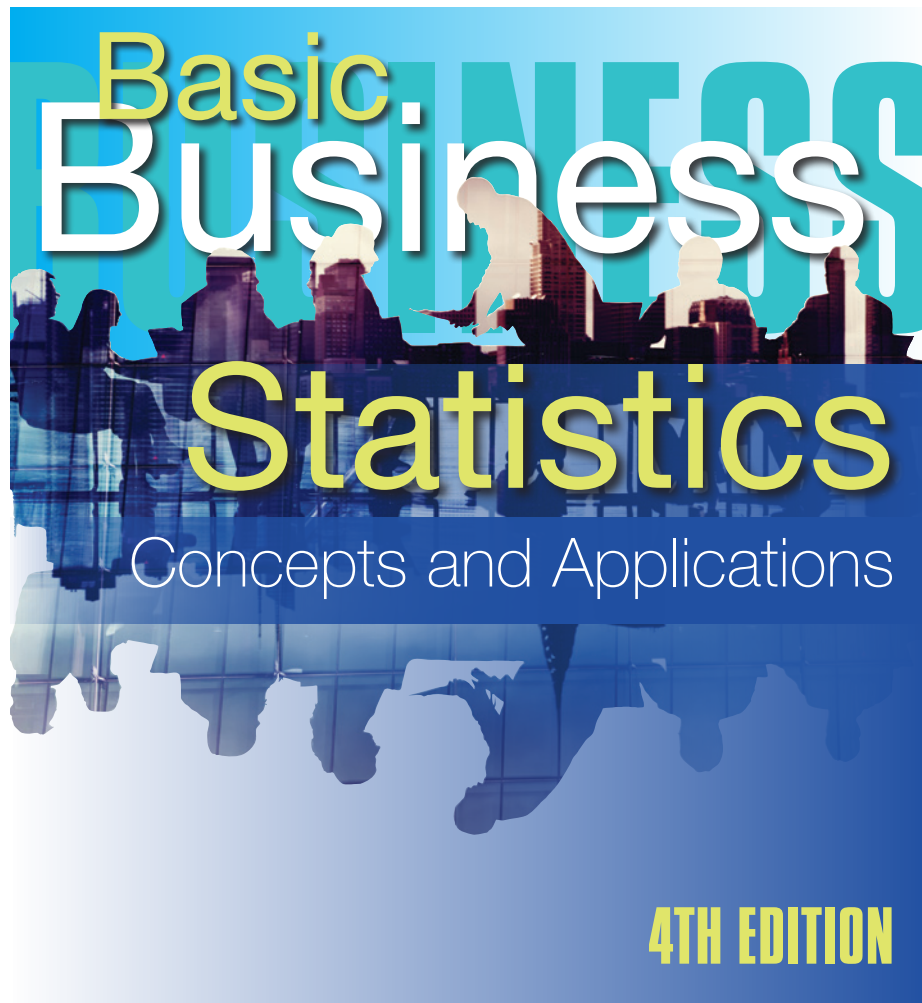
ALWAYS LEARNING

PEARSON

Basic
Business
Statistics

Concepts and Applications

4TH EDITION



Berenson Levine Szabat O'Brien Jayne Watson

Copyright © Pearson Australia (a division of Pearson Australia Group Pty Ltd) 2016

Pearson Australia
707 Collins Street
Melbourne VIC 3008

www.pearson.com.au

Authorised adaptation from the United States edition entitled *Basic Business Statistics*, 13th edition, ISBN 0321870026 by Berenson, Mark L.; Levine, David M.; Szabat, Kathryn A., published by Pearson Education, Inc., Copyright © 2015.

Fourth adaptation edition published by Pearson Australia Group Pty Ltd, Copyright © 2016

The *Copyright Act 1968* of Australia allows a maximum of one chapter or 10% of this book, whichever is the greater, to be copied by any educational institution for its educational purposes provided that that educational institution (or the body that administers it) has given a remuneration notice to Copyright Agency Limited (CAL) under the Act. For details of the CAL licence for educational institutions contact: Copyright Agency Limited, telephone: (02) 9394 7600, email: info@copyright.com.au

All rights reserved. Except under the conditions described in the *Copyright Act 1968* of Australia and subsequent amendments, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

Senior Learning Architect: Joanne Hobson
Development Editor: Anna Carter
Senior Project Manager: Sandra Goodall
Digital Product Developer: Erin Nixon
Digital Content Developer: Udarie Fernando
Copyright and Pictures Editor: Kim Morgan
Copy Editor: Julie Ganner
Proofreader: Katie Millar
Indexer: Garry Cousins
Cover and internal design by Liz Nicholson, design BITE Pty Ltd
Cover photograph © Raw Pixel/Shutterstock.com
Typeset by iEnergizer Aptara[®], Ltd

Printed in China

1 2 3 4 5 20 19 18 17 16

National Library of Australia
Cataloguing-in-Publication Data

Creator: Berenson, Mark L., author.
Title: Basic business statistics : concepts and applications /
Mark L. Berenson, David M. Levine, Kathryn A. Szabat,
Martin O'Brien, Nicola Jayne, Judith Watson.

Edition: 4th edition
ISBN: 9781486018956 (paperback)
ISBN: 9781486019410 (Vital Source)

Notes: Includes index.
Subjects: Commercial statistics.
Business mathematics.
Statistics.

Other Creators/Contributors:
Levine, David M., 1946– author.
Szabat, Kathryn A., author.
O'Brien, Martin, author.
Jayne, Nicola, author.
Watson, Judith, 1956– author.

Dewey Number: 519.5

Every effort has been made to trace and acknowledge copyright. However, should any infringement have occurred, the publishers tender their apologies and invite copyright owners to contact them.

brief contents

	Preface	x
	Acknowledgements	xi
	How to use this book	xii
	About the authors	xvii
PART 1	PRESENTING AND DESCRIBING INFORMATION	
	1 Introduction and data collection	2
	2 Presenting data in tables and charts	20
	3 Numerical descriptive measures	58
PART 2	MEASURING UNCERTAINTY	
	4 Basic probability	111
	5 Some important discrete probability distributions	141
	6 The normal distribution and other continuous distributions	171
	7 Sampling distributions	204
PART 3	DRAWING CONCLUSIONS ABOUT POPULATIONS BASED ONLY ON SAMPLE INFORMATION	
	8 Confidence interval estimation	243
	9 Fundamentals of hypothesis testing: One-sample tests	278
	10 Hypothesis testing: Two-sample tests	320
	11 Analysis of variance	360
PART 4	DETERMINING CAUSE AND MAKING RELIABLE FORECASTS	
	12 Simple linear regression	410
	13 Introduction to multiple regression	459
	14 Time-series forecasting and index numbers	499
	15 Chi-square tests	564
PART 5	FURTHER TOPICS IN STATS	
	16 Multiple regression model building	608
	17 Decision making	635
	18 Statistical applications in quality and productivity management	659
	19 Further non-parametric tests	693
	Appendices A to F	A-1
	Glossary	G-1
	Index	I-1

detailed contents

PART 1

PRESENTING AND DESCRIBING INFORMATION

1	Introduction and data collection	2
1.1	Basic concepts of statistics	3
1.2	The growth of statistics and information technology	5
1.3	Collecting data	6
1.4	Types of variables	8
	<i>Summary</i>	11
	<i>Key terms</i>	12
	<i>Chapter review problems</i>	12
	<i>Appendix 1</i> Introduction to using Microsoft Windows and Excel	13
2	Presenting data in tables and charts	20
2.1	Tables and charts for categorical data	21
2.2	Organising numerical data	26
2.3	Tables and graphs for numerical data	29
2.4	Cross-tabulations	38
2.5	Scatter diagrams and time-series plots	42
2.6	Misusing graphs and ethical issues	45
	<i>Summary</i>	49
	<i>Key terms</i>	50
	<i>Chapter review problems</i>	50
	<i>Appendix 2</i> Using statistical software for tables and charts	53
3	Numerical descriptive measures	58
3.1	Measures of central tendency, variation and shape	59
3.2	Numerical descriptive measures for a population	80
3.3	Calculating numerical descriptive measures from a frequency distribution	85
3.4	Five-number summary and box-and-whisker plots	87
3.5	Covariance and the coefficient of correlation	90
3.6	Pitfalls in numerical descriptive measures and ethical issues	96
	<i>Summary</i>	97
	<i>Key formulas</i>	97

<i>Key terms</i>	99
<i>Chapter review problems</i>	99
<i>Appendix 3</i> Using statistical software for descriptive statistics	101
End of Part 1 problems	103

PART 2

MEASURING UNCERTAINTY

4	Basic probability	111
4.1	Basic probability concepts	112
4.2	Conditional probability	120
4.3	Bayes' theorem	127
4.4	Counting rules	132
4.5	Ethical issues and probability	136
	<i>Summary</i>	137
	<i>Key formulas</i>	137
	<i>Key terms</i>	137
	<i>Chapter review problems</i>	138
	<i>Appendix 4</i> Using statistical software for basic probability	140
5	Some important discrete probability distributions	141
5.1	Probability distribution for a discrete random variable	142
5.2	Covariance and its application in finance	146
5.3	Binomial distribution	150
5.4	Poisson distribution	157
5.5	Hypergeometric distribution	161
5.6	(<i>Online Topic</i>) Using the poisson distribution to approximate the binomial distribution	164
	<i>Summary</i>	165
	<i>Key formulas</i>	165
	<i>Key terms</i>	166
	<i>Chapter review problems</i>	166
	<i>Appendix 5</i> Using statistical software for discrete probability distributions	168
6	The normal distribution and other continuous distributions	171
6.1	Continuous probability distributions	172
6.2	The normal distribution	173

6.3	Evaluating normality	188		
6.4	The uniform distribution	192		
6.5	The exponential distribution	194		
6.6	The normal approximation to the binomial distribution	197		
	<i>Summary</i>	200		
	<i>Key formulas</i>	200		
	<i>Key terms</i>	200		
	<i>Chapter review problems</i>	201		
	<i>Appendix 6</i> Using statistical software for continuous probability distributions	202		
7	Sampling distributions	204		
7.1	Sampling distributions	205		
7.2	Sampling distribution of the mean	205		
7.3	Sampling distribution of the proportion	215		
7.4	Types of survey sampling methods	218		
7.5	Evaluating survey worthiness	223		
	<i>Summary</i>	227		
	<i>Key formulas</i>	227		
	<i>Key terms</i>	228		
	<i>Chapter review problems</i>	228		
	<i>Appendix 7</i> Using statistical software for sampling distributions	230		
	End of Part 2 problems	232		
PART 3				
DRAWING CONCLUSIONS ABOUT POPULATIONS BASED ONLY ON SAMPLE INFORMATION				
8	Confidence interval estimation	243		
8.1	Confidence interval estimation for the mean (σ known)	244		
8.2	Confidence interval estimation for the mean (σ unknown)	249		
8.3	Confidence interval estimation for the proportion	255		
8.4	Determining sample size	258		
8.5	Applications of confidence interval estimation in auditing	264		
8.6	More on confidence interval estimation and ethical issues	271		
	<i>Summary</i>	272		
	<i>Key formulas</i>	272		
	<i>Key terms</i>	272		
	<i>Chapter review problems</i>	273		
	<i>Appendix 8</i> Using statistical software for confidence intervals and sample size determination	277		
9	Fundamentals of hypothesis testing: One-sample tests	278		
9.1	Hypothesis-testing methodology	279		
9.2	Z test of hypothesis for the mean (σ known)	285		
9.3	One-tail tests	292		
9.4	t test of hypothesis for the mean (σ unknown)	297		
9.5	Z test of hypothesis for the proportion	303		
9.6	The power of a test	307		
9.7	Potential hypothesis-testing pitfalls and ethical issues	312		
	<i>Summary</i>	315		
	<i>Key formulas</i>	316		
	<i>Key terms</i>	316		
	<i>Chapter review problems</i>	316		
	<i>Appendix 9</i> Using statistical software for one-sample tests of hypothesis	319		
10	Hypothesis testing: Two-sample tests	320		
10.1	Comparing the means of two independent populations	321		
10.2	Comparing the means of two related populations	333		
10.3	F test for the difference between two variances	340		
10.4	Comparing two population proportions	347		
	<i>Summary</i>	352		
	<i>Key formulas</i>	353		
	<i>Key terms</i>	354		
	<i>Chapter review problems</i>	354		
	<i>Appendix 10</i> Using statistical software for two-sample tests	358		
11	Analysis of variance	360		
11.1	The completely randomised design: One-way analysis of variance	361		
11.2	The randomised block design	375		
11.3	The factorial design: Two-way analysis of variance	384		
	<i>Summary</i>	396		
	<i>Key formulas</i>	397		
	<i>Key terms</i>	398		
	<i>Chapter review problems</i>	398		
	<i>Appendix 11</i> Using statistical software for ANOVA	401		
	End of Part 3 problems	403		

PART 4**DETERMINING CAUSE AND MAKING RELIABLE FORECASTS****12 Simple linear regression 410**

- 12.1 Types of regression models 411
- 12.2 Determining the simple linear regression equation 413
- 12.3 Measures of variation 422
- 12.4 Assumptions 428
- 12.5 Residual analysis 428
- 12.6 Measuring autocorrelation: The Durbin–Watson statistic 432
- 12.7 Inferences about the slope and correlation coefficient 437
- 12.8 Estimation of mean values and prediction of individual values 444
- 12.9 Pitfalls in regression and ethical issues 449

Summary 452

Key formulas 452

Key terms 453

Chapter review problems 453

Appendix 12 Using statistical software for simple linear regression 457

13 Introduction to multiple regression 459

- 13.1 Developing the multiple regression model 460
- 13.2 R^2 , adjusted R^2 and the overall F test 466
- 13.3 Residual analysis for the multiple regression model 469
- 13.4 Inferences concerning the population regression coefficients 471
- 13.5 Testing portions of the multiple regression model 475
- 13.6 Using dummy variables and interaction terms in regression models 480
- 13.7 Collinearity 490

Summary 492

Key formulas 492

Key terms 493

Chapter review problems 493

Appendix 13 Using statistical software for multiple regression 497

14 Time-series forecasting and index numbers 499

- 14.1 The importance of business forecasting 500

- 14.2 Component factors of the classical multiplicative time-series model 501
- 14.3 Smoothing the annual time series 503
- 14.4 Least-squares trend-fitting and forecasting 510
- 14.5 The Holt–Winters method for trend-fitting and forecasting 522
- 14.6 Autoregressive modelling for trend-fitting and forecasting 525
- 14.7 Choosing an appropriate forecasting model 534
- 14.8 Time-series forecasting of seasonal data 539
- 14.9 Index numbers 547
- 14.10 Pitfalls in time-series forecasting 555

Summary 556

Key formulas 556

Key terms 557

Chapter review problems 558

Appendix 14 Using statistical software for time-series forecasting and index numbers 560

15 Chi-square tests 564

- 15.1 Chi-square test for the difference between two proportions (independent samples) 565
- 15.2 Chi-square test for differences between more than two proportions 572
- 15.3 Chi-square test of independence 579
- 15.4 Chi-square goodness-of-fit tests 584
- 15.5 Chi-square test for a variance or standard deviation 589

Summary 592

Key formulas 592

Key terms 593

Chapter review problems 593

Appendix 15 Using statistical software for chi-square tests 597

End of Part 4 problems 598

PART 5**FURTHER TOPICS IN STATS****16 Multiple regression model building 608**

- 16.1 Quadratic regression model 609
- 16.2 Using transformations in regression models 615
- 16.3 Influence analysis 618
- 16.4 Model building 620
- 16.5 Pitfalls in multiple regression and ethical issues 630

<i>Summary</i>	632	<i>Chapter review problems</i>	689
<i>Key formulas</i>	632	<i>Appendix 18</i> Using statistical software for statistical applications in quality management	691
<i>Key terms</i>	632		
<i>Chapter review problems</i>	633		
<i>Appendix 16</i> Software for multiple regression model building	634		
17 Decision making	635	19 Further non-parametric tests	693
17.1 Payoff tables and decision trees	636	19.1 McNemar test for the difference between two proportions (related samples)	694
17.2 Criteria for decision making	640	19.2 Wilcoxon rank sum test: Non-parametric analysis for two independent populations	697
17.3 Decision making with sample information	649	19.3 Wilcoxon signed ranks test: Non-parametric analysis for two related populations	703
17.4 Utility	654	19.4 Kruskal–Wallis rank test: Non-parametric analysis for the one-way ANOVA	707
<i>Summary</i>	656	19.5 Friedman rank test: Non-parametric analysis for the randomised block design	711
<i>Key formulas</i>	656	<i>Key formulas</i>	715
<i>Key terms</i>	656	<i>Summary</i>	715
<i>Chapter review problems</i>	656	<i>Key terms</i>	716
<i>Appendix 17</i> Using statistical software for decision making	658	<i>Chapter review problems</i>	716
		<i>Appendix 19</i> Using statistical software for non-parametric tests	717
18 Statistical applications in quality and productivity management	659	End of Part 5 problems	718
18.1 Total quality management	660		
18.2 Six sigma management	662	Appendices	
18.3 The theory of control charts	663	A Review of arithmetic, algebra and logarithms	A-2
18.4 Control chart for the proportion – the p chart	665	B Summation notation	A-4
18.5 The red bead experiment: Understanding process variability	671	C Statistical symbols and Greek alphabet	A-7
18.6 Control chart for an area of opportunity – the c chart	673	D PHStat2 user's guide	A-8
18.7 Control charts for the range and the mean	676	E Tables	A-17
18.8 Process capability	682	F Using Microsoft Excel with this text	A-44
<i>Summary</i>	688	Glossary	G-1
<i>Key formulas</i>	688	Index	I-1
<i>Key terms</i>	689		

preface

This fourth Australasian and Pacific edition of *Basic Business Statistics: Concepts and Applications* continues to build on the strengths of the third edition, and extends the outstanding teaching foundation of the previous American editions, authored by Berenson, Levine and Szabat.

The teaching philosophy of this text is based upon the principles of the American book, but each chapter has once again been carefully revised to include practical examples and a language and style that is more applicable to Australasian and Pacific readers.

In preparation for this edition we again asked lecturers from around the country to comment on the format and content of the third edition and, based on those comments, the authors have worked to create a text that is more accessible – but no less authoritative – for students.

Part 5 contains additional chapters: Chapter 16 on multiple regression and model building, Chapter 17 on decision making, Chapter 18 on statistical applications in quality and productivity management, and Chapter 19 on further non-parametric tests. These chapters can be found within the MyLab and student download page via our catalogue.

The new edition has continued with a ‘real-world’ focus, to take students beyond the pure theory. Some chapters have a completely new opening scenario, focusing on a person or company, which serves to introduce key concepts covered in the chapter. The scenario is interwoven throughout the chapter to reinforce the concepts to the student. Multiple in-chapter examples have been updated that highlight real Australasian and Pacific data.

Vivien Spyraakis has come on board this edition and played a vital part in assisting the authors with their research, as well as conducting interviews with the new Real people, real stats.

The **Real people, real stats** feature that opens each of the text’s five parts is composed of a personal interview highlighting how *real* people in *real* business situations apply the principles of statistics to their jobs. The interviewees include:

- Part 1** David McCourt *BDO*
- Part 2** Ellouise Roberts *Deloitte Access Economics*
- Part 3** Rod Battye *Tourism Research Australia*
- Part 4** Gautam Gangopadhyay *Endeavour Energy*
- Part 5** Deborah O’Mara *The University of Sydney*

Judith Watson
Nicola Jayne
Martin O’Brien

acknowledgements

When developing the new edition of *Basic Business Statistics*, we were mindful of retaining the strengths of the current edition, but also of the need to build on those strengths, to enhance the text and to ensure wider reader appeal and useability.

We are indebted to the following academics who contributed to the new edition.

Researcher

Vivien Spyrakis *Previous experience as a casual lecturer and tutor at the University of NSW and The University of Sydney; currently working as a private tutor in the area of advanced maths and statistics*

Reviewers

Jill Wright *Monash University*

Renato Andrin Villano *UNE Business School, University of New England Australia*

Robert Brooks *Monash Business School*

Michael Steele *Universiti Brunei Darussalam*

Dr Matthew J. Beck *University of Sydney*

Ranjith Ihalanayake *College of Business, Victoria University*

Dr Woon Weng Wong *RMIT University*

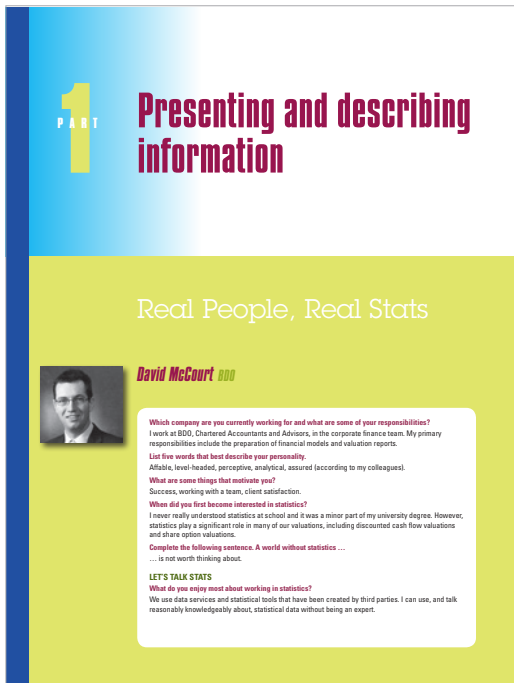
Kuldeep Kumar *Bond University*

Dr Somkiat Kitjongthawonkul *Australian Catholic University*

Jo-ann Larkins *Monash University*

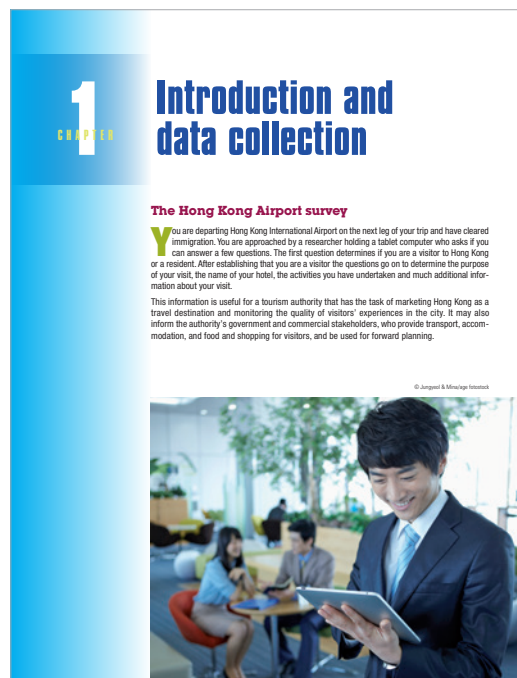
The enormous task of writing a book of this scope was possible only with the expert assistance of all these friends and colleagues and that of the editorial and production staff at Pearson Australia. We gratefully acknowledge their invaluable contributions at every stage of this project, collectively and, now, individually. We thank the following people at Pearson Australia: Joanne Hobson, Senior Learning Architect; Anna Carter, Development Editor; Sandra Goodall, Senior Project Manager; Julie Ganner, Copy Editor; and Kim Morgan, Copyright and Pictures Editor.

how to use this book

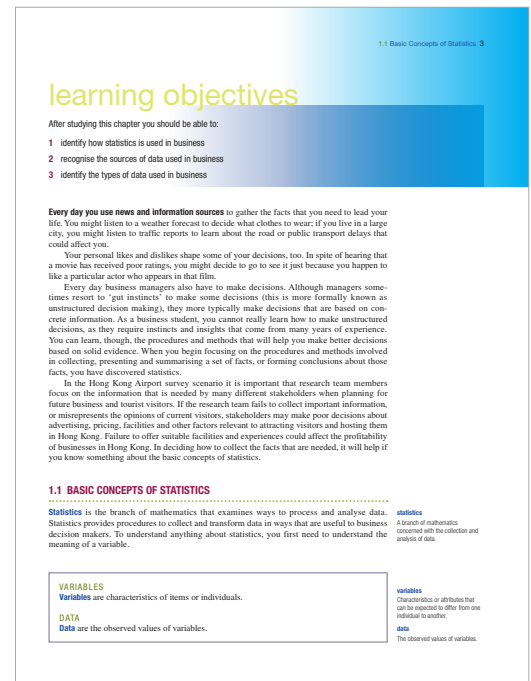


Real people, real stats interviews open each part. These introduce *real people* working in *real* business environments, using statistics to tackle *real* business challenges.

Chapter-opening scenarios show how statistics are used in everyday life. The scenarios introduce the concepts to be covered, showing the relevance of using particular statistical techniques. The problem is woven throughout each chapter, showing the connection between statistics and their use in business, as well as keeping you motivated.

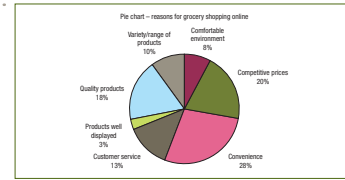


Learning objectives introduce you to the key concepts to be covered in each chapter, and are signposted in the margins where they are covered within the chapter.



What type of chart should you use? The selection of a chart depends on your intention. If a comparison of categories is most important, use a bar chart. If observing the portion of the whole that lies in a particular category is most important, use a pie chart.

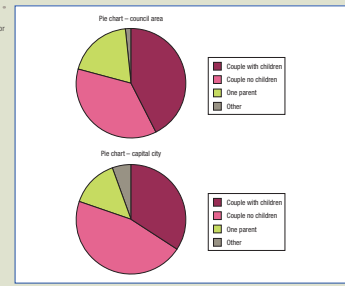
Figure 2.3 Microsoft Excel pie chart of the reasons for grocery shopping online



EXAMPLE 2.3 PIE CHART FOR FAMILY TYPE

The council and capital city demographics are presented as summary tables. Use the summary tables given for family type to construct and interpret pie charts for the capital city and the council area.

Figure 2.4 Microsoft Excel pie chart for family type



Real world, business examples are included throughout the chapter. These are designed to show the multiple applications of statistics, while helping you to learn the statistics techniques.

Emphasis on data output and interpretation

The authors believe that the use of computer software is an integral part of learning statistics. Our focus emphasises analysing data by interpreting the output from Microsoft Excel while reducing emphasis on doing calculations. Excel 2013 changes to statistical functions are reflected in the operations shown in this edition.

In the coverage of hypothesis testing in Chapters 9 to 11, extensive computer output is included so that the focus can be placed on the *p*-value approach. In our coverage of simple linear regression in Chapter 12, we assume that a software program will be used and our focus is on interpretation of the output, not on hand calculations.

Summaries are provided at the end of each chapter, to help you review the key content.

Key terms are signposted in the margins when they are first introduced, and are referenced to page numbers at the end of each chapter, helping you to revise key terms and concepts for the chapter.

End-of-section problems are divided into *Learning the basics* and *Applying the concepts*.

End-of-part problems challenge the student to make decisions about the appropriate technique to apply, to carry out that technique and to interpret the data meaningfully.*

Australasian and Pacific data sets are used for the problems in each chapter. These files are contained on the Pearson website.

Ethical issues sections are integrated into many chapters, raising issues for ethical consideration.

658 CHAPTER SEVENTEEN Decision Making

17 Assess your progress

Summary

In this chapter you learned how to develop payoff tables and decision trees, to use various criteria to choose between alternative courses of action and to revise probabilities in light of sample information using Bayes' theorem. In the scenario at the beginning of the chapter, you learned how a manager of a managed fund could use these tools to decide whether to purchase shares in company A or to purchase them in company B. You found that company B's shares had a higher expected monetary value, a lower expected opportunity loss but a lower return-to-risk ratio.

Key formulas

Expected monetary value
 $EMV(j) = \sum_{i=1}^n x_{ij} p_i$ (17.1)

Expected opportunity loss
 $EOL(j) = \sum_{i=1}^n L_{ij} p_i$ (17.2)

Expected value of perfect information
 $EVPI = \text{expected profit under certainty} - \text{expected monetary value of the best alternative}$ (17.3)

Return-to-risk ratio
 $RTRR(j) = \frac{EMV(j)}{\sigma_j}$ (17.4)

Key terms

alternative courses of action	603	expected profit under certainty	610	return-to-risk ratio (RTRR)	612
decision criteria	603	expected value of perfect information	604	risk-avertor's curve	622
decision tree	604	(EVP)	610	risk-neutral curve	622
events or states of the world	603	opportunity loss	605	risk-seeker's curve	622
expected monetary value (EMV)	608	payoff	603	utility	621
expected opportunity loss (EOL)	609	payoff table	603		

Chapter review problems

CHECKING YOUR UNDERSTANDING

17.25 What is the difference between an event and an alternative course of action?
 17.26 What are the advantages and disadvantages of a payoff table as compared with a decision tree?
 17.27 How are opportunity losses calculated from payoffs?
 17.28 Why can't an opportunity loss be negative?
 17.29 How does expected monetary value (EMV) differ from expected opportunity loss (EOL)?
 17.30 What is the meaning of the expected value of perfect information (EVPI)?

17.31 How does the expected value of perfect information differ from the expected profit under certainty?
 17.32 What are the advantages and disadvantages of using expected monetary value (EMV) as compared with the return-to-risk ratio (RTRR)?
 17.33 How is Bayes' theorem used to revise probabilities in light of the sample information?
 17.34 What is the difference between a risk averter and a risk seeker?
 17.35 Why should you use utilities instead of payoffs in certain circumstances?

End of Part 1 problems

A.1 A sample of 500 shoppers was selected in a large metropolitan area to obtain consumer behaviour information. Among the questions asked was, 'Do you enjoy shopping for clothing?' The results are summarised in the following cross-classification table.

Enjoy shopping for clothing	Gender		Total
	Male	Female	
Yes	136	224	360
No	104	36	140
Total	240	260	500

a. Construct contingency tables based on total percentages, row percentages and column percentages.
 b. Construct a side-by-side bar chart of enjoy shopping for clothing based on gender.
 c. What conclusions do you draw from these analyses? One of the major measures of the quality of service provided by any organisation is the speed with which the organisation responds to customer complaints. A large family-held department store selling furniture and flooring, including carpet, has undergone major expansion in the past few years. In particular, the flooring department has expanded from two installation crews to an installation supervisor, a messenger and 15 installation crews. During a recent year the company got 50 complaints about carpet installation. The following data represent the number of days between receipt of the complaint and resolution of the complaint.

<PROBABILITY>

54	5	36	127	31	27	12	2	123	81	74	27
11	126	110	29	61	35	94	31	26	5		
12	4	165	32	29	28	26	25	1	14	13	
13	10	5	27	4	62	30	22	36	26	20	23
33	68										

a. Construct frequency and percentage distributions.
 b. Construct histogram and percentage polygons.
 c. Construct a cumulative percentage distribution and plot the corresponding ogive.
 d. Calculate the mean, median, first quartile and third quartile.
 e. Calculate the variance, standard deviation, range, interquartile range and coefficient of variation.
 f. Construct and interpret a box-and-whisker plot.
 g. What conclusions can you reach concerning the magnesium content of this day's production?
A.5 The Achilles organisation releases the results of recent surveys at its website <www.achilles.com>. Go to this site and search for **global online survey**. Select an article of interest to you – for example, Global Online Shopping Habits.
 a. Give an example of a categorical variable found in the presentation.
 b. Give an example of a numerical variable found in the presentation.
 c. Is the variable you selected in (b) discrete or continuous?

A.2 For each fund calculate the geometric rate of return for these years and for five years.
A.3 The annual crediting rates (after tax and fees) on several managed superannuation investment funds are:

Superannuation fund	2013	2012	2011	2010	2009
Conservative	8.8	4.9	8.0	8.8	-2.5
Balanced	13.1	2.5	10.5	9.1	-7.6
Growth	15.7	0.7	11.7	10.3	-12.8
High growth	19.7	-2.4	13.7	9.3	-16.8

*The solutions are calculated using the (raw) Excel output. If you use the rounded figures presented in the text to reproduce these answers there may be minor differences.

MyMathLab Global

a guided tour for students and educators

Study Plan

A study plan is generated from each student's results on a pre-test. Students can clearly see which topics they have mastered and, more importantly, which they need to work on.

The screenshot shows the 'Study Plan' page for a course. The left sidebar contains navigation options: Course Home, Course Materials (Study Plan Builder, Study Plan, Assignments, eText, Multimedia, Student Resources), Communication, Results, Product Support, and Educator Tools. The main content area is titled 'Study Plan Recommendations' and shows 'You have earned 0 of 72 mastery points (MP)'. It lists sections to practice and master with 'Practise' and 'Quiz Me' buttons and mastery point counts.

Section	Practise	Quiz Me	Mastery Points
1.4 Types of variables	Practise	Quiz Me	0 of 1 MP
2.1 Tables and charts for categorical data	Practise	Quiz Me	0 of 1 MP
2.2 Organising numerical data	Practise	Quiz Me	0 of 1 MP
2.3 Tables and charts for numerical data	Practise	Quiz Me	0 of 1 MP
2.4 Cross-tabulations	Practise	Quiz Me	0 of 1 MP

Unlimited Practice

Each MyMathLab Global comes with preloaded assignments, including select end-of-chapter questions, all of which are automatically graded. Many study plan and educator-assigned exercises contain algorithmically generated values to ensure students get as much practice as they need.

As students work through study plan or homework exercises, instant feedback and tutorial resources guide them towards understanding.

The screenshot shows the 'Study Plan Builder' page. The left sidebar is the same as the previous screenshot. The main content area is titled 'Homework and Tests: Quizzes & Tests' and shows a table of 'All Quizzes and Tests' and 'Sample Tests and Quizzes'.

Due	Assignment	Time Limit	Attempts	Gradebook Score
	Chapter 1 Study Plan Builder		0 of ∞	

Sample Tests and Quizzes	Attempts	Gradebook Score
Chapter 2 Study Plan Builder	0 of ∞	
Chapter 3 Study Plan Builder	0 of ∞	
Chapter 4 Study Plan Builder	0 of ∞	
Chapter 5 Study Plan Builder	0 of ∞	
Chapter 6 Study Plan Builder	0 of ∞	
Chapter 7 Study Plan Builder	0 of ∞	
Chapter 8 Study Plan Builder	0 of ∞	
Chapter 9 Study Plan Builder	0 of ∞	
Chapter 10 Study Plan Builder	0 of ∞	
Chapter 11 Study Plan Builder	1 of ∞	see score
Chapter 12 Study Plan Builder	0 of ∞	
Chapter 13 Study Plan Builder	0 of ∞	
Chapter 14 Study Plan Builder	0 of ∞	
Chapter 15 Study Plan Builder	0 of ∞	

4.1 Basic probability concepts Overview

Problem 4.1

0 correct | 0 of 5 complete

Five coins are tossed.

a. Give an example of a simple event.
 b. Give an example of a joint event.
 c. What is the complement of heads on the first coin?

a. Which of the following is a simple event?

A. Getting heads on the first coin and heads on the fourth coin
 B. Getting heads on the fifth coin
 C. Getting tails on all coins
 D. Getting tails on no coins

Click to select your answer, then click Check Answer.

2 parts remaining

Clear All Check Answer Close

[Help Me Solve This](#)
[View an Example](#)
[Textbook](#)
[StatCrunch](#)
[Ask My Instructor](#)
[Print](#)

Learning Resources

To further reinforce understanding, study plan and homework problems link to the following learning resources:

- eText linked to sections for all study plan questions
- Help Me Solve This, which walks students through the problem with step-by-step help and feedback without giving away the answer
- StatCrunch.

Courses Hello, EDU Mastercourse Account Help & Support Sign Out

Berenson Course Sample MyMathLab Global

course settings

modify Multimedia

▼ Course Home
 Assignment Calendar
 Purchase Options

▼ Course Materials
 Study Plan Builder
 Study Plan
 Assignments
 eText
 Multimedia
 Student Resources

► Communication
 Results
 Product Support
 ► Educator Tools

Multimedia Library


Berenson, Basic Business Statistics, 4e (Aust)

To view the multimedia resources available for your textbook, make your selection(s) below.

Chapter: 1. Introduction and data collection
 Section: All Sections
 Media Type: Select All Video

Find Now

https://mediaplayer.pearsoncmg.com/media_dark_clean_set.width.640__set.height.480__aw/aw_mml_shared_1/statistics/StatTalkVideo...
 https://mediaplayer.pearsoncmg.com/media_dark_clean_set.width.640__set.height.480__aw/aw_mml_shared_1/statistics/StatTalkVideo...



StatTalk Videos

Fun-loving statistician Andrew Vickers takes to the streets of Brooklyn, New York to demonstrate important statistical concepts through interesting stories and real-life events. This series of videos and corresponding auto-graded questions will help students to understand statistics.

EDUCATOR RESOURCES

A suite of resources provided to assist with delivery of the text, as well as to support teaching and learning.

Solutions Manual

The Solutions Manual provides educators with detailed, accuracy-verified solutions to all the in-chapter and end-of-chapter problems in the book.

Test Bank

The Test Bank provides a wealth of accuracy-verified testing material. Updated for the new edition, each chapter offers a wide variety of true/false and multiple-choice questions, arranged by learning objective and tagged by AACSB standards.

Every question is available in TestGen® software for both Windows® and Macintosh® computers. This easy-to-use testing software is a valuable test preparation tool that allows educators to customise, save and generate classroom tests.

PowerPoint lecture slides

A comprehensive set of PowerPoint slides can be used by educators for class presentations or by students for lecture preview or review. They include key figures and tables, as well as a summary of key concepts and examples from the text.

Digital image PowerPoint slides

All the diagrams and tables from the text are available for lecturer use.

about the authors

Judith Watson

Judith Watson teaches in the Business School at UNSW Australia. She has extensive experience in lecturing and administering undergraduate and post-graduate Quantitative Methods courses.

Judith's keen interest in student support led her to establish the Peer Assisted Support Scheme (PASS) in 1996 and she has coordinated this program for many years. She served as her faculty's academic adviser from 2001 to 2004. Judith has been the recipient of a number of awards for teaching. She received the inaugural Australian School of Business Outstanding Teaching Innovations Award in 2008 and the 2012 Bill Birkett Award for Teaching Excellence. She also won the UNSW Vice Chancellor's Award for Teaching Excellence in 2012 and a Citation of Outstanding Contributions to Student Learning from the Australian Government's Office for Learning. Judith is interested in using online learning technology to engage students and has created a number of adaptive e-learning tutorials for mathematics and statistics and cartoon-style videos to explain statistical concepts.



Dr Nicola Jayne

Nicola Jayne is a lecturer in the Southern Cross Business School at the Lismore campus of Southern Cross University. She has been teaching quantitative units since being appointed to the university in 1993 after several years at Massey University in New Zealand. Nicola has lectured extensively in Business and Financial Mathematics, Discrete Mathematics and Statistics, both undergraduate and postgraduate, as well as various Pure Mathematics units.

Her academic qualifications from Massey University include a Bachelor of Science (majors in Mathematics and Statistics), a Bachelor of Science with Honours (first class) and a Doctor of Philosophy, both in Mathematics. Nicola also has a Graduate Certificate in Higher Education (Learning & Teaching) from Southern Cross University.

She was the recipient of a Vice Chancellor's Citation for an Outstanding Contribution to Student Learning in 2011.



Dr Martin O'Brien

Dr Martin O'Brien is a senior lecturer and Head of the Discipline of Economics at the School of Economics, University of Wollongong. Martin earned his Bachelor of Commerce (first-class honours) and PhD in Economics at the University of Newcastle. Martin's PhD and subsequent published research is in the general area of labour economics, and specifically the exploration of older workers' labour force participation in Australia in the context of an ageing society. He has taught a wide range of quantitative subjects at university level, including business statistics, quantitative analysis for decision making, econometrics, financial modelling, business research methods and quality management. Martin also has a keen interest in the development of new teaching technologies and the analysis of alternative teaching methods such as practice-into-theory.



the originating authors

Mark L. Berenson is Professor of Management and Information Systems at Montclair State University (Montclair, New Jersey) and also Professor Emeritus of Statistics and Computer Information Systems at Bernard M. Baruch College (City University of New York). He currently teaches graduate and undergraduate courses in statistics and in operations management in the School of Business, and an undergraduate course in international justice and human rights that he co-developed in the College of Humanities and Social Sciences.

Berenson received a BA in economic statistics and an MBA in business statistics from City College of New York and a PhD in business from the City University of New York. Berenson's research has been published in *Decision Sciences Journal of Innovative Education*, *Review of Business Research*, *The American Statistician*, *Communications in Statistics*, *Psychometrika*, *Educational and Psychological Measurement*, *Journal of Management Sciences and Applied Cybernetics*, *Research Quarterly*, *Stats Magazine*, *The New York Statistician*, *Journal of Health Administration Education*, *Journal of Behavioral Medicine* and *Journal of Surgical Oncology*. His invited articles have appeared in *The Encyclopedia of Measurement & Statistics* and *Encyclopedia of Statistical Sciences*. He is co-author of 11 statistics texts published by Prentice Hall, including *Statistics for Managers Using Microsoft Excel*, *Basic Business Statistics: Concepts and Applications* and *Business Statistics: A First Course*.

Over the years, Berenson has received several awards for teaching and for innovative contributions to statistics education. In 2005, he was the first recipient of the Catherine A. Becker Service for Educational Excellence Award at Montclair State University, and in 2012 he was the recipient of the Khubani/Telebrands Faculty Research Fellowship in the School of Business.

David M. Levine is Professor Emeritus of Statistics and Computer Information Systems at Baruch College (City University of New York). He received BBA and MBA degrees in statistics from City College of New York and a PhD from New York University in industrial engineering and operations research. He is recognised in the United States as a leading innovator in statistics education and is the co-author of 14 books, including such best-selling statistics textbooks as *Statistics for Managers Using Microsoft Excel*, *Basic Business Statistics: Concepts and Applications*, *Business Statistics: A First Course* and *Applied Statistics for Engineers and Scientists Using Microsoft Excel and Minitab*.

Levine is also the co-author of *Even You Can Learn Statistics: A Guide for Everyone Who Has Ever Been Afraid of Statistics*, currently in its second edition, *Six Sigma for Green Belts and Champions* and *Design for Six Sigma for Green Belts and Champions*; and is the author of *Statistics for Six Sigma Green Belts*, all published by FT Press, a Pearson imprint; and *Quality Management*, third edition, published by McGraw-Hill/Irwin. He is also the author of *Video Review of Statistics* and *Video Review of Probability*, both published by Video Aided Instruction, and the statistics module of the MBA primer published by Cengage Learning. He has published articles in various journals, including *Psychometrika*, *The American Statistician*, *Communications in Statistics*, *Decision Sciences Journal of Innovative Education*, *Multivariate Behavioral*

Research, *Journal of Systems Management*, *Quality Progress* and *The American Anthropologist*, and he has given numerous talks at the Decision Sciences Institute (DSI), American Statistical Association (ASA), and Making Statistics More Effective in Schools and Business (MSMESB) conferences. Levine has also received several awards for outstanding teaching and curriculum development from Baruch College.

Kathryn A. Szabat is Associate Professor and Chair of Business Systems and Analytics at LaSalle University. She teaches undergraduate and graduate courses in business statistics and operations management.

Szabat's research has been published in *International Journal of Applied Decision Sciences*, *Accounting Education*, *Journal of Applied Business and Economics*, *Journal of Healthcare Management* and *Journal of Management Studies*. Scholarly chapters have appeared in *Managing Adaptability, Intervention, and People in Enterprise Information Systems*; *Managing, Trade, Economies and International Business*; *Encyclopedia of Statistics in Behavioral Science*; and *Statistical Methods in Longitudinal Research*.

Szabat has provided statistical advice to numerous business, non-business and academic communities. Her more recent involvement has been in the areas of education, medicine and non-profit capacity building.

Szabat received a BS in mathematics from State University of New York at Albany and MS and PhD degrees in statistics, with a cognate in operations research, from the Wharton School of the University of Pennsylvania.

PART 1

Presenting and describing information

Real People, Real Stats



David McCourt BDO

Which company are you currently working for and what are some of your responsibilities?

I work at BDO, Chartered Accountants and Advisors, in the corporate finance team. My primary responsibilities include the preparation of financial models and valuation reports.

List five words that best describe your personality.

Affable, level-headed, perceptive, analytical, assured (according to my colleagues).

What are some things that motivate you?

Success, working with a team, client satisfaction.

When did you first become interested in statistics?

I never really understood statistics at school and it was a minor part of my university degree. However, statistics play a significant role in many of our valuations, including discounted cash flow valuations and share option valuations.

Complete the following sentence. A world without statistics ...

... is not worth thinking about.

LET'S TALK STATS

What do you enjoy most about working in statistics?

We use data services and statistical tools that have been created by third parties. I can use, and talk reasonably knowledgeably about, statistical data without being an expert.

Describe your first statistics-related job or work experience. Was this a positive or a negative experience?

The first time I can recall using statistics was for a share option valuation. We had to determine the share price volatility based on historical share price data. There are about half a dozen methods that can be used, all with various advantages and disadvantages. I did and still find this analysis interesting.

What do you feel is the most common misconception about your work held by students who are studying statistics? Please explain.

Statistics provides information to support our analysis and decisions. However, the information is never perfect, and subjectivity and commercial common sense play a large part in our work.

Do you need to be good at maths to understand and use statistics successfully?

I think you need to have a logical and well-structured approach to problems. These skills would probably make you good at both maths and statistics.

Is there a high demand for statisticians in your industry (or in other industries)? Please explain.

The finance industry is heavily reliant on statistics. I expect there is high demand for statisticians from the various data providers, and in a number of specialist areas (e.g. insurance).

PRESENTING AND DESCRIBING INFORMATION

Does data collection play an important role in the decisions you make for your business/work? Please explain.

Accurate data collection is essential to our valuation projects. Although our work involves a degree of commercial acumen, it is essential that the data supports and justifies these decisions. We also aggregate data for internal business use to measure staff productivity, business performance and forecasting budgets.

Describe a project that you have worked on recently that might have involved data collection. Please be specific.

We recently valued an infrastructure asset using the discounted cash flow model. The model requires two essential inputs: the forecast of future cash flows of the asset, and the discount rate that reflects the riskiness of those cash flows. To arrive at an appropriate discount rate we generally analyse comparable companies for an indication of the level of risk that should be attributed to the asset to be valued. In this exercise there are several instances of data collection. We collect five-year historical stock data for numerous comparable companies as an

initial indication of risk. We then collect data on key financial indicators to assess the degree of comparability between the stock and the asset to be valued. To determine the risk-free rate and the market-risk premium, 10-year government bond rate data is collected.

How are these data usually summarised? What are some positives and negatives of these summary techniques?

We generally organise the collected data into Microsoft Excel workbooks. The main advantage of using this software is the ease of data analysis. Some powerful data analysis tools include data tables, What-If Analysis, Solver, charting and common statistical functions. Some shortcomings we have encountered using Excel is that data sometimes need to be rearranged depending on the analysis, [there can be] problems with inconsistent or missing data, and output can sometimes be incomplete. These factors increase the likelihood of errors in data analysis; however, for the purposes of corporate finance, Excel is generally sufficient as a means of summarising and analysing the data collected.

In your experience, what is the most commonly referred to measure of central tendency? What benefits does this measure offer over others?

In valuations, we generally prefer to use the median as a measure of central tendency rather than mean or mode. We find that the mean has one main disadvantage: it is particularly susceptible to outliers. When looking at comparable companies there are often outliers caused by one-off business issues that are irrelevant for the purposes of comparing our business. We very rarely use mode given that it only really coincides with the central tendency of data where the distribution is centre-heavy and there are generally few recurring figures in the data set.

Why is it important to be aware of the spread/variation of data points in a sample? What are the consequences of not knowing this type of information about your sample?

Without an understanding of the spread and variation of a data set there is no context to the measure of central tendency applied. A measure of central tendency summarises the data into a single value while the spread and variation of data gives an indication of how reliable an average or median summary of collected data is. For example, if the spread of values in the data set is relatively large it suggests the mean is not as representative, and a smoothing of data is required, when compared to a data set with a smaller range. Adopting a mean without reference to the spread can taint our analysis and results in a lack of validity to our decisions that are based on the data.

1 CHAPTER

Introduction and data collection

The Hong Kong Airport survey

You are departing Hong Kong International Airport on the next leg of your trip and have cleared immigration. You are approached by a researcher holding a tablet computer who asks if you can answer a few questions. The first question determines if you are a visitor to Hong Kong or a resident. After establishing that you are a visitor the questions go on to determine the purpose of your visit, the name of your hotel, the activities you have undertaken and much additional information about your visit.

This information is useful for a tourism authority that has the task of marketing Hong Kong as a travel destination and monitoring the quality of visitors' experiences in the city. It may also inform the authority's government and commercial stakeholders, who provide transport, accommodation, and food and shopping for visitors, and be used for forward planning.

© Jungyeol & Mina/age fotostock



learning objectives

After studying this chapter you should be able to:

- 1 identify how statistics is used in business
- 2 recognise the sources of data used in business
- 3 identify the types of data used in business

Every day you use news and information sources to gather the facts that you need to lead your life. You might listen to a weather forecast to decide what clothes to wear; if you live in a large city, you might listen to traffic reports to learn about the road or public transport delays that could affect you.

Your personal likes and dislikes shape some of your decisions, too. In spite of hearing that a movie has received poor ratings, you might decide to go to see it just because you happen to like a particular actor who appears in that film.

Every day business managers also have to make decisions. Although managers sometimes resort to ‘gut instincts’ to make some decisions (this is more formally known as unstructured decision making), they more typically make decisions that are based on concrete information. As a business student, you cannot really learn how to make unstructured decisions, as they require instincts and insights that come from many years of experience. You can learn, though, the procedures and methods that will help you make better decisions based on solid evidence. When you begin focusing on the procedures and methods involved in collecting, presenting and summarising a set of facts, or forming conclusions about those facts, you have discovered statistics.

In the Hong Kong Airport survey scenario it is important that research team members focus on the information that is needed by many different stakeholders when planning for future business and tourist visitors. If the research team fails to collect important information, or misrepresents the opinions of current visitors, stakeholders may make poor decisions about advertising, pricing, facilities and other factors relevant to attracting visitors and hosting them in Hong Kong. Failure to offer suitable facilities and experiences could affect the profitability of businesses in Hong Kong. In deciding how to collect the facts that are needed, it will help if you know something about the basic concepts of statistics.

1.1 BASIC CONCEPTS OF STATISTICS

Statistics is the branch of mathematics that examines ways to process and analyse data. Statistics provides procedures to collect and transform data in ways that are useful to business decision makers. To understand anything about statistics, you first need to understand the meaning of a variable.

statistics

A branch of mathematics concerned with the collection and analysis of data.

VARIABLES

Variables are characteristics of items or individuals.

DATA

Data are the observed values of variables.

variables

Characteristics or attributes that can be expected to differ from one individual to another.

data

The observed values of variables.

Examples of variables are your gender, your major field of study, the amount of money you have in your wallet or purse, and the amount of time you spend in the shower each day. The key aspect of the word *variable* is the idea that items differ and people differ. The person next to you may be male rather than female, may be majoring in a field of study different from your own, almost certainly has a different amount of money in their wallet, and may be less conscious of saving water than you are. You should distinguish between a variable, such as gender or cost, and its *value* for an individual observation (e.g. ‘male’ or ‘\$55’). Collections of values are known as *data*.

operational definition

Defines how a variable is to be measured.

All variables should have an **operational definition** – a universally accepted meaning that is clear to all associated with an analysis. Without operational definitions, confusion can occur. An example of a situation where operational definitions are needed is for the process of data gathering by the Australian Bureau of Statistics. The ABS needs to collect information about groups of people living together, and so it defines a ‘household’ as ‘one or more persons, at least one of whom is at least 15 years of age, usually resident in the same private dwelling’ whereas a ‘family’ has the operational definition of ‘two or more persons, one of whom is at least 15 years of age, who are related by blood, marriage (registered or de facto), adoption, step or fostering, and who are usually resident in the same household’ (Australian Bureau of Statistics, 2010, *Standards for income variables*, Cat. No. 1287.0).

Comparing these definitions we can see that it is possible for a household to contain more than one family. The household is the preferred statistical unit for income analysis.

Now that the terms *variables* and *data* have been defined, you need to understand the meaning of the terms *population*, *sample*, *parameter* and *statistic*.

population

A collection of all members of a group being investigated.

sample

The portion of the population selected for analysis.

parameter

A numerical measure of some population characteristic.

statistic

A numerical measure that describes a characteristic of a sample.

POPULATION

A **population** consists of all the members of a group about which you want to draw a conclusion.

SAMPLE

A **sample** is the portion of the population selected for analysis.

PARAMETER

A **parameter** is a numerical measure that describes a characteristic of a population.

STATISTIC

A **statistic** is a numerical measure that describes a characteristic of a sample.

Examples of populations are all the full-time students at a university, all the registered voters in New Zealand and all the people who were customers of the local shopping centre last weekend. The term *population* is not limited to groups of people. We could refer to a population of all motor vehicles registered in Victoria. Two factors need to be specified when defining a population:

1. the entity (e.g. people or motor vehicles)
2. the boundary (e.g. registered to vote in New Zealand or registered in Victoria for road use).

Samples could be selected from each of the populations mentioned above. Examples include 10 full-time students selected for a focus group, 500 registered voters in New Zealand who were contacted by telephone for a political poll, 30 customers at the shopping centre who were asked to complete a market research survey and all the vehicles registered in Victoria that are more than 10 years old. In each case, the people or the vehicles in the sample represent a portion, or subset, of the people or vehicles comprising the population.

The average amount spent by all the customers at the local shopping centre last weekend is an example of a *parameter*. Information from all the shoppers in the entire population is needed to calculate this parameter.

The average amount spent by the 30 customers completing the market research survey is an example of a *statistic*. Information from a sample of only 30 of the shopping centre's customers is used in calculating the statistic.

Statistics, itself, is divided into two branches, both of which are applicable to managing a business. **Descriptive statistics** focuses on collecting, summarising and presenting a set of data. **Inferential statistics** uses sample data to draw conclusions about a population.

Descriptive statistics has its roots in the record-keeping needs of large political and social organisations. Refining the methods of descriptive statistics is an ongoing task for government statistical agencies such as the Australian Bureau of Statistics and Statistics New Zealand as they prepare for each Census. In Australia, a Census is scheduled to be carried out every five years (e.g. 2011 and 2016) to count the entire population and to collect data about education, occupation, languages spoken and many other characteristics of the citizens. A large amount of planning and training is necessary to ensure that the data collected represent an accurate record of the population's characteristics at the Census date. However, despite the best planning, such an immense data collection task can be affected by external factors. The New Zealand Census originally planned for 8 March 2011 had to be cancelled due to chaos caused by the major earthquake on 22 February in Christchurch. A new Census was delayed until March 2013.

The foundation of inferential statistics is based on the mathematics of probability theory. Inferential methods use sample data to calculate statistics that provide estimates of the characteristics of the entire population.

Today, applications of statistical methods can be found in different areas of business. Accounting uses statistical methods to select samples for auditing purposes and to understand the cost drivers in cost accounting. Finance uses statistical methods to choose between alternative portfolio investments and to track trends in financial measures over time. Management uses statistical methods to improve the quality of the products manufactured or the services delivered by an organisation. Marketing uses statistical methods to estimate the proportion of customers who prefer one product over another and to draw conclusions about what advertising strategy might be most useful in increasing sales of a product.

1.2 THE GROWTH OF STATISTICS AND INFORMATION TECHNOLOGY

During the past century, statistics has played an important role in spurring the use of information technology and, in turn, such technology has spurred the wider use of statistics. At the beginning of the twentieth century, the expanding data-handling requirements associated with the United States federal Census led directly to the development of tabulating machines that were the forerunners of today's business computer systems. Statisticians such as Pearson, Fisher, Gosset, Neyman, Wald and Tukey established the techniques of modern inferential statistics as an alternative to analysing large sets of population data that had become increasingly costly, time-consuming and cumbersome to collect. The development of early computer systems permitted others to develop computer programs to ease the calculation and data-processing burdens imposed by those techniques. Over time, greater use of statistical methods by business decision makers and advances in computer capacity have led to the development of even more sophisticated statistical methods.

Today, when you hear of retailers investing in a 'customer-relationship management system', or CRM, or a packaged goods producer engaging in 'data mining' to uncover consumer preferences, you should realise that statistical techniques form the foundations of such cutting-edge applications of information technology. As global information storage increases dramatically, businesses are rapidly coming to terms with how to analyse **big data** – data sets so large and varied that conventional software cannot readily handle them. (Think of the huge volume of data produced each day by people using Visa, Facebook, eBay and Twitter.) Even though

descriptive statistics

The field that focuses on summarising or characterising a set of data.

inferential statistics

Uses information from a sample to draw conclusions about a population.

big data

Large data sets characterised by their volume, velocity and variety.

statistical packages

Computer programs designed to perform statistical analysis.

cutting-edge applications might require custom programming, for many years businesses have had access to **statistical packages** such as Minitab, SPSS/PASW Statistics, SAS and Stata – standardised sets of programs that help managers use a wide range of statistical techniques by automating the data processing and calculations these techniques require.

The leasing and training costs associated with statistical packages have led many to consider using some of the graphical and statistical functions of Microsoft Excel. However, you need to be aware that many statisticians have concerns about the accuracy and completeness of the statistical results produced by early versions of Excel. Invalid results could be produced, especially when the data sets were very large or had unusual statistical properties (see reference 1). Microsoft Excel 2010 and subsequent versions made some significant improvements in statistical functions (see references 2 and 3) but it would still be wise to be careful about the data and the analysis you are undertaking.

1.3 COLLECTING DATA

LEARNING OBJECTIVE

1

Identify how statistics is used in business

In the Hong Kong Airport scenario, identifying the data that need to be collected is an important step in the process of marketing the city and operational planning. Some of the data will come from consumers through market research. It is important that the correct inferences are drawn from the research and that appropriate statistical methods assist planners and designers to make the right decisions.

Managing a business effectively requires collecting the appropriate data. In most cases, the data are measurements acquired from items in a sample. The samples are chosen from populations in such a manner that the sample is as representative of the population as possible. The most common technique to ensure proper representation is to use a random sample. (See Chapter 7 for a detailed discussion of sampling techniques.)

Many different types of circumstances require the collection of data:

- A marketing research analyst needs to assess the effectiveness of a new television advertisement.
- A pharmaceutical manufacturer needs to determine whether a new drug is more effective than those currently in use.
- An operations manager wants to monitor a manufacturing process to find out whether the quality of product being produced is conforming to company standards.
- An auditor wants to review the financial transactions of a company to determine whether or not the company is in compliance with generally accepted accounting principles.
- A potential investor wants to determine which firms within which industries are likely to have accelerated growth in a period of economic recovery.

Identifying Sources of Data

Identifying the most appropriate source of data is a critical aspect of statistical analysis. If biases, ambiguities or other types of errors flaw the data being collected, even the most sophisticated statistical methods will not produce accurate information. Four important sources of data are:

- data distributed by an organisation or an individual
- a designed experiment
- a survey
- an observational study.

Data sources are classified as either **primary sources** or **secondary sources**. When the data collector is the one using the data for analysis, the source is primary. When another organisation or individual has collected the data that are used for analysis by an organisation or individual, the source is secondary.

Organisations and individuals that collect and publish data typically use this information as a primary source and then let others use the data as a secondary source. For example, the

LEARNING OBJECTIVE

2

Recognise sources of data used in business

primary sources

Provide information collected by the data analyst.

secondary sources

Provide data collected by another person or organisation.

Australian federal government collects and distributes data in this way for both public and private purposes. The Australian Bureau of Statistics oversees a variety of ongoing data collection in areas such as population, the labour force, energy, and the environment and health care, and publishes statistical reports. The Reserve Bank of Australia collects and publishes data on exchange rates, interest rates and ATM and credit card transactions.

Market research firms and trade associations also distribute data pertaining to specific industries or markets. Investment services such as Aspect Huntley provide financial data on a company-by-company basis. Syndicated services such as ACNielsen provide clients with data enabling the comparison of client products with those of their competitors. Daily newspapers in print and online formats are filled with numerical information about share prices, weather conditions and sports statistics.

As listed above, conducting an experiment is another important data-collection source. For example, to test the effectiveness of laundry detergent, an experimenter determines which brands in the study are more effective in cleaning soiled clothes by actually washing dirty laundry instead of asking customers which brand they believe to be more effective. Proper experimental designs are usually the subject matter of more advanced texts, because they often involve sophisticated statistical procedures. However, some fundamental experimental design concepts are considered in Chapter 11.

Conducting a survey is a third important data source. Here, the people being surveyed are asked questions about their beliefs, attitudes, behaviours and other characteristics. Responses are then edited, coded and tabulated for analysis.

The Hong Kong Airport survey

Travellers in the departure lounge of the busy Hong Kong International Airport are asked to complete a survey with questions about various aspects of their visit to the city and future travel plans. The interviewer first asks if the traveller is a resident or a visitor. If the traveller is a visitor the survey proceeds. The survey includes these questions:

- How many visits have you made to Hong Kong prior to this one? _____
- How long is it since your visit here? _____
- How satisfied were you with your accommodation?
 Very satisfied Satisfied Undecided Dissatisfied Very dissatisfied
- How many times during this visit did you travel by ferry? _____
- Shopping in Hong Kong stores gives good value for money
 Almost always Very infrequently
 Sometimes Never
- Was the purpose of your visit business? Yes No
- Are you likely to return to Hong Kong in the next 12 months? Yes No

You have been asked to review the survey. What type of data does the survey seek to collect?

What type of information can be generated from the data of the completed survey? How can the research company's clients use this information when planning for future visitors? What other questions would you suggest for the survey?

focus group

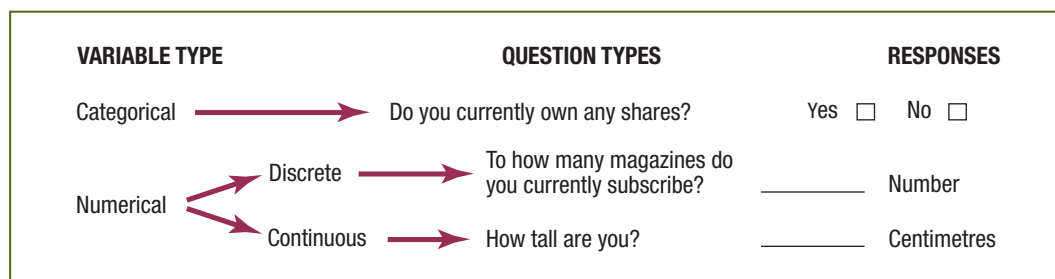
A group of people who are asked about attitudes and opinions for qualitative research.

Conducting an observational study is the fourth important data source. In such a study, a researcher observes the behaviour directly, usually in its natural setting. Observational studies take many forms in business. One example is the **focus group**, a market research tool that is used to elicit unstructured responses to open-ended questions. In a focus group, a moderator leads the discussion and all the participants respond to the questions asked. Other, more structured types of studies involve group dynamics and consensus building and use various organisational-behaviour tools such as brainstorming, the Delphi technique and the nominal-group method. Observational study techniques are also used in situations in which enhancing teamwork or improving the quality of products and services are management goals.

1.4 TYPES OF VARIABLES

As illustrated in Figure 1.1, there are two types of variables – categorical and numerical, sometimes referred to as qualitative and quantitative variables respectively.

Figure 1.1
Types of variables



categorical variables

Take values that fall into one or more categories.

Categorical variables yield categorical responses, such as *yes* or *no* or *male* or *female* answers. An example is the response to the question ‘Do you currently own any shares?’ because it is limited to a simple *yes* or *no* answer. Another example is the response to the question in the Hong Kong Airport survey (presented on page 7), ‘Are you likely to return to Hong Kong in the next 12 months?’ Categorical variables can also yield more than one possible response. For example, ‘On which days of the week are you most likely to use public transport?’

numerical variables

Take numbers as their observed responses.

Numerical variables yield numerical responses, such as your height in centimetres. Other examples are ‘How many times during this visit did you travel by ferry?’ (from the Hong Kong Airport survey) or the response to the question ‘To how many magazines do you currently subscribe?’

LEARNING OBJECTIVE 3

Identify the types of data used in business

discrete variables

Can only take a finite or countable number of values.

There are two types of numerical variables: discrete and continuous. **Discrete variables** produce numerical responses that arise from a counting process. ‘The number of magazines subscribed to’ is an example of a discrete numerical variable because the response is one of a finite number of integers. You subscribe to zero, one, two and so on magazines.

Continuous variables produce numerical responses that arise from a measuring process.

Your height is an example of a continuous numerical variable because the response takes on any value within a continuum or interval, depending on the precision of the measuring instrument. For example, your height may be 158 cm, 158.3 cm or 158.2945 cm, depending on the precision of the available instruments.

continuous variables

Can take any value between specified limits.

No two people are exactly the same height and the more precise the measuring device used, the greater the likelihood of detecting differences in their heights. However, most measuring devices are not sophisticated enough to detect small differences. Hence, *tied observations* are often found in experimental or survey data even though the variable is truly continuous and, theoretically, all values of a continuous variable are different.

Levels of Measurement and Types of Measurement Scales

Data are also described in terms of their level of measurement. There are four widely recognised levels of measurement: nominal, ordinal, interval and ratio scales.

Nominal and ordinal scales Data from a categorical variable are measured on a nominal scale or on an ordinal scale. A **nominal scale** (Figure 1.2) classifies data into various distinct categories in which no ranking is implied. In the Hong Kong Airport survey, the answer to the question ‘Are you likely to return to Hong Kong in the next 12 months?’ is an example of a nominally scaled variable, as is your favourite soft drink, your political party affiliation and your gender. Nominal scaling is the weakest form of measurement because you cannot specify any ranking across the various categories.

nominal scale

A classification of categorical data that implies no ranking.

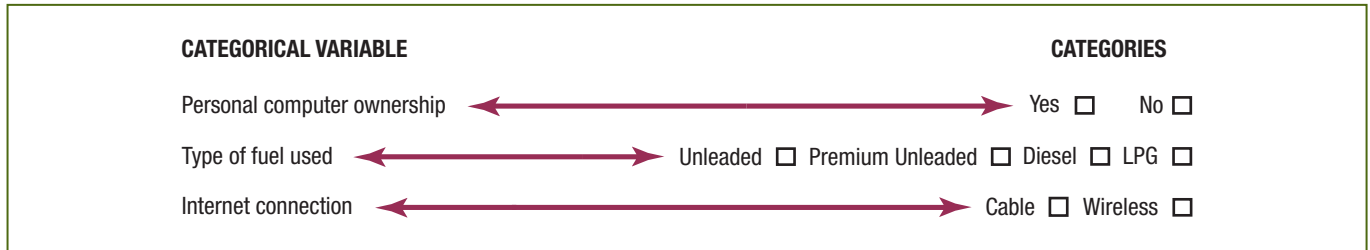


Figure 1.2 Examples of nominal scaling

An **ordinal scale** classifies data into distinct categories in which ranking is implied. In the Hong Kong Airport survey, the answers to the question ‘Shopping in Hong Kong stores gives good value for money’ represent an ordinal scaled variable because the responses ‘almost always, sometimes, very infrequently and never’ are ranked in order of frequency. Figure 1.3 lists other examples of ordinal scaled variables.

ordinal scale

Scale of measurement where values are assigned by ranking.

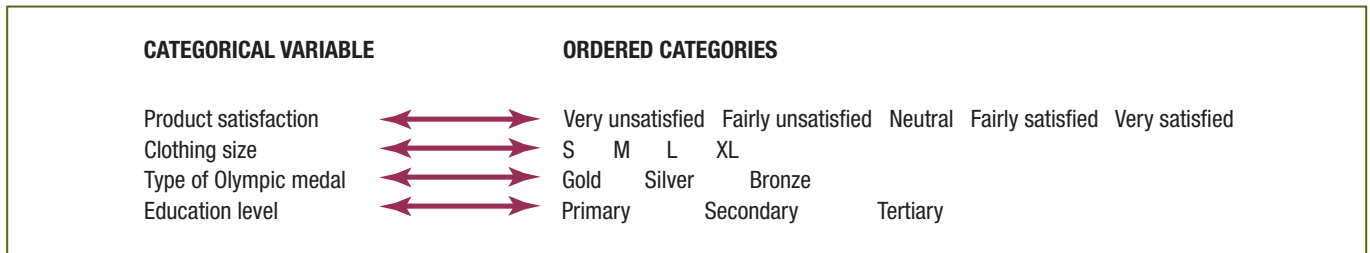


Figure 1.3 Examples of ordinal scaling

Ordinal scaling is a stronger form of measurement than nominal scaling because an observed value classified into one category possesses more or less of a property than does an observed value classified into another category. However, ordinal scaling is still a relatively weak form of measurement because the scale does not account for the amount of the differences *between* the categories. The ordering implies only *which* category is ‘greater’, ‘better’ or ‘more preferred’ – not by *how much*.

Interval and ratio scales Data from a numerical variable are measured on an interval or ratio scale. An **interval scale** (Figure 1.4) is an ordered scale in which the difference between measurements is a meaningful quantity but does not involve a true zero point. For example, sports

interval scale

A ranking of numerical data where differences are meaningful but there is no true zero point.

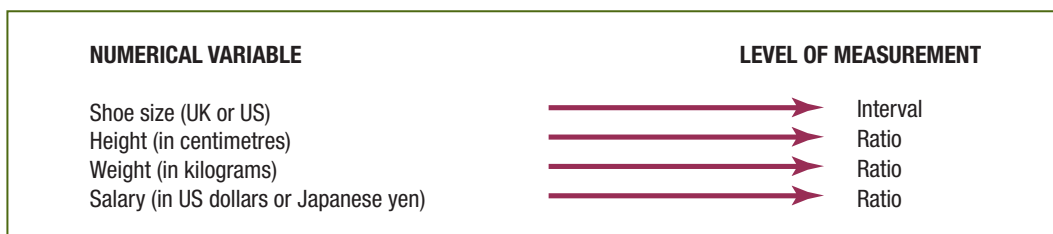


Figure 1.4

Examples of interval and ratio scales

shoes for adults are often sold in Australia marked with sizes based on the US or UK system. Neither system has a true zero size. The size below an adult size 1 is a child's size 13. However, in each system the intervals between sizes are equal.

ratio scale

A ranking where the differences between measurements involve a true zero point.

A **ratio scale** is an ordered scale in which the difference between the measurements involves a true zero point, as in length, weight, age or salary measurements, and the ratio of two values is meaningful. In the Hong Kong Airport survey, the number of times a visitor travelled by ferry is an example of a ratio scaled variable. As another example, a carton that weighs 40 kg is twice as heavy as one that weighs 20 kg.

Data measured on an interval scale or on a ratio scale constitute the highest levels of measurement. They are stronger forms of measurement than an ordinal scale, because you can determine not only which observed value is the largest but also by how much. Interval and ratio scales may apply for either discrete or continuous data.

think
about
this

Telephone polling

Companies such as Newspoll regularly undertake market research and political polling conducted by phone interviews. A phone poll conducted by Newspoll in Sydney in November 2014 asked questions about a number of topics. Some were demographic questions about the number of people who lived in the household and the age, income, occupation and marital status of the participant. What would be the purpose of asking such questions?

The other questions could be divided into three sections. The first section related to voting intentions for the next state election and the level of satisfaction with the premier and the opposition leader. The second section asked the participant's opinion on the renewal of the federal government's ban on super trawlers. The third section asked a number of questions about domestic and international air travel undertaken in the past year. These questions covered areas such as the purpose of travel, the airlines used and level of satisfaction.

Who would use the data collected in this poll? If you were designing a similar poll, how would you construct questions to collect data for the variables referred to above?

After answering these questions, see how Newspoll reports opinion poll results by looking at <www.newspoll.com.au>.

Problems for Section 1.4

LEARNING THE BASICS

- 1.1 Three different types of drinks are sold at a fast-food restaurant – soft drinks, fruit juices and coffee.
 - a. Explain why the type of drinks sold is an example of a categorical variable.
 - b. Explain why the type of drinks sold is an example of a nominally scaled variable.
- 1.2 Soft drinks are sold in three sizes in a fast-food restaurant – small, medium and large. Explain why the size of the soft drink is an example of an ordinal scaled variable.
- 1.3 Suppose that you measure the time it takes to download an MP3 file from the Internet.
 - a. Explain why the download time is a numerical variable.
 - b. Explain why the download time is a ratio scaled variable.

APPLYING THE CONCEPTS

- 1.4** For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.
- Number of mobile phones per household
 - Length (in minutes) of the longest mobile call made per month
 - Whether all mobile phones in the household use the same telecommunications provider
 - Whether there is a landline telephone in the household
- 1.5** The following information is collected from students as they leave the campus bookshop during the first week of classes:
- Amount of time spent shopping in the bookshop
 - Number of textbooks purchased
 - Name of degree
 - Gender
- Classify each of these variables as categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.
- 1.6** For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the level of measurement.
- Name of Internet provider
 - Amount of time spent surfing the Internet per week
 - Number of emails received in a week
 - Number of online purchases made in a month
- 1.7** Suppose the following information is collected from Andrew and Fiona Chen on their application for a home loan mortgage at Metro Home Loans:
- Monthly expenses: \$2,056
 - Number of dependants being supported by applicant(s): 2
 - Annual family salary income: \$105,000
 - Marital status: Married
- Classify each of the responses by type of data and level of measurement.
- 1.8** One of the variables most often included in surveys is income. Sometimes the question is phrased, 'What is your income (in thousands of dollars)?' In other surveys, the respondent is asked to 'Place an X in the circle corresponding to your income group' and given a number of ranges to choose from.
- In the first format, explain why income might be considered either discrete or continuous.
 - Which of these two formats would you prefer to use if you were conducting a survey? Why?
 - Which of these two formats would probably bring you a greater rate of response? Why?
- 1.9** The director of research at the e-business section of a major department store wants to conduct a survey throughout a Australia to determine the amount of time working women spend shopping online for clothing in a typical month.
- Describe the population and the sample of interest, and indicate the type of data the director might wish to collect.
 - Develop a first draft of the questionnaire needed in (a) by writing a series of three categorical questions and three numerical questions that you feel would be appropriate for this survey.
- 1.10** A university researcher designs an experiment to see how generous participants will be in giving to charity. Discuss how the experiment might give different results from a survey of the same subjects about donations to charity.
- 1.11** Before a company undertakes an online marketing campaign it needs to consider information about its own current sales and the sales made by its competitors. What sources of data might it use?

Assess your progress

1

Summary

In this chapter you have studied data collection and the various types of data used in business. In the Hong Kong International Airport scenario you were asked to review the visitor survey which will be

used to provide information to the tourism authority planning staff (see page 7). Three of the questions shown will produce numerical data and four will produce categorical data. The responses to the

first question (number of previous visits to Hong Kong) are discrete, and the responses to the second question (length of time since last visit) are continuous. After the data have been collected, they must

be organised and prepared in order to make various analyses. The next two chapters develop tables and charts and a variety of descriptive numerical measures that are useful for data analysis.

Key terms

big data	5	interval scale	9	ratio scale	10
categorical variables	8	nominal scale	9	sample	4
continuous variables	8	numerical variables	8	secondary sources	6
data	3	operational definition	4	statistic	4
descriptive statistics	5	ordinal scale	9	statistics	3
discrete variables	8	parameter	4	statistical packages	6
focus group	8	population	4	variables	3
inferential statistics	5	primary sources	6		

Chapter review problems

CHECKING YOUR UNDERSTANDING

- 1.12** What is the difference between a sample and a population?
- 1.13** What is the difference between a statistic and a parameter?
- 1.14** What is the difference between descriptive and inferential statistics?
- 1.15** What is the difference between a categorical and a numerical variable?
- 1.16** What is the difference between a discrete and a continuous variable?
- 1.17** What is an operational definition and why is it so important?
- 1.18** What are the four types of measurement scales?

APPLYING THE CONCEPTS

- 1.19** The Australasian Data and Story Library OZDASL <www.statsci.org/data/index.html> is an online library of data files and stories that illustrate the use of basic statistical methods. The stories are classified by method and by topic. Go to this site and click on 'First course in statistics'. Pick a story and summarise how statistics were used in the story.
- 1.20** Make a list of six ways you have used or encountered statistics in the past week. Think about what you read or heard in a news report or saw on a commercial website. Also think whether you made a bet or participated in a survey.
- 1.21** The Australian Bureau of Statistics <www.abs.gov.au> site contains survey information on people, business, geography and other topics. Go to the site and find the latest version of *Labour Force, Australia* (Cat. No. 6202.0).
- Briefly describe the Labour Force survey.
 - Give an example of a categorical variable found in this survey.
 - Give an example of a numerical variable found in this survey.
 - Is the variable you selected in (c) discrete or continuous?
- 1.22** The Australian Bureau of Statistics website allows users to access a large amount of Census data online. Go to <www.abs.gov.au/websitedbs/censushome.nsf/home/Data> and click on the latest Census year in the section titled 'Find Census data by location'. Select a local area and then view the QuickStats.

- Give an example of a categorical variable found in this summary of survey results.
 - Give an example of a numerical variable found in this summary of survey results.
 - Is the variable you selected in (b) discrete or continuous?
- 1.23** Detailed information on airport and airline on-time performance can be found at <www.flightstats.com/go/Airport/airportScorecard.do>. Select an airport name or code, either local or international, and the latest date then click on Search. Explore the arrivals and departures data in the four panels that are given.
- Which of the four types of data sources listed in Section 1.3 do you think were used here?
 - Name a categorical variable for which observations were collected.
 - Name a numerical variable for which observations were collected.
- 1.24** From June 2013 the NRMA, a major Australian motoring organisation, conducted a web-based survey that sought to check the visibility of school safety zone signage throughout NSW and the ACT. An invitation to participate via closed-link survey was sent to school principals and P&Cs and an open-link survey was available to the general public (<www.mynrma.com.au/media/School_Zones_report_2013_NRMA.pdf>).
- Describe the population(s) for this survey.
 - Describe the sample(s) for this survey.
 - Can you identify potential difficulties in amalgamating data from the two sources?
- 1.25** A manufacturer of flavoured milk is planning to survey households in Tasmania to determine the purchasing habits of consumers. Among the questions to be included are those that relate to:
- where flavoured milk is primarily purchased
 - what flavour of milk is purchased most often
 - how many people living in the household drink flavoured milk
 - the total number of millilitres of flavoured milk drunk in the past week by members of the household
- Describe the population.

- b. For each of the four items listed, indicate whether the variable is categorical or numerical. If numerical, is it discrete or continuous?
- c. Develop five categorical questions for the survey.
- d. Develop five numerical questions for the survey.
- 1.26** A new bus network is proposed for a north-eastern Sydney region. A survey is sent out to residents asking questions which relate to:
1. the resident's age
 2. frequency of bus use
 3. usual ticket type purchased
 4. main purpose of using the bus
- a. Describe the population.
- b. Indicate whether each of the questions above is categorical or numerical.
- c. Develop two more numerical questions and state whether the variables are discrete or continuous.
- d. Develop two more categorical questions.

References

1. McCullough, B. D. & B. Wilson, 'On the accuracy of statistical procedures in Microsoft Excel 97', *Computational Statistics and Data Analysis*, 31 (1999): 27–37.
2. Microsoft Corporation at <<http://office.microsoft.com/en-au/excel-help/what-s-new-changes-made-to-excel-functions-HA010355760.aspx>>, accessed December 2013.
3. Microsoft Corporation at <<http://office.microsoft.com/en-001/excel-help/new-functions-in-excel-2013-HA103980604.aspx>>, accessed December 2013.

Appendix 1

Introduction to using Microsoft Windows and Excel

Read 'Using Windows' if you are unfamiliar with the basics of Microsoft Windows operations that are required in order to use Microsoft Excel. Read the sections on using Microsoft Excel on a PC or a Mac commencing on page 16 if you plan to use Microsoft Excel with this text.

Note: Throughout this appendix and all other appendices in this text, the symbol → is used to represent a sequence of menu selections. For example, the instruction 'select **File** → **Open**' means first select the **File** menu choice and then select the **Open** choice from the submenu that appears. Also note that the methods described below may need to be varied depending on your version of Windows and device. For instance, the way you access the Start menu on a tablet will be different depending on whether or not you have a keyboard with mousepad attached.

Using Windows on a PC

Using the mouse

In Microsoft Windows you frequently use a mouse, touchpad or other pointing device to make choices by pointing to

an onscreen object and pressing a mouse button. Windows expects pointing devices to have two buttons, one called the primary button, the other the secondary button. You can move your mouse and press and release the mouse buttons in the following ways:

- **Click or Select** Move the mouse pointer over an object and press the primary button.
- **Drag** Move the mouse pointer over an object. Then, while pressing and holding down the primary button, move the mouse pointer somewhere else on the screen and release the primary button. Dragging either moves objects to another part of the screen or allows you to select multiple items.
- **Double-click** Move the mouse pointer over an object and click the primary button twice in rapid succession.
- **Right-click** Move the mouse pointer over an object and click the secondary button.
- **Copy and paste** There are several ways to copy and paste. Perhaps the most simple is using the mouse to highlight, then Ctrl+c for copying and Ctrl+v for pasting in the new position.

Microsoft Windows defines the left mouse button as the primary button and the right button as the secondary button (this gives rise to the phrase ‘right-click’), but you can swap the definitions by selecting the Windows Mouse Control Panel icon.

Opening programs

There are two ways to open a program such as Microsoft Excel for use:

- **Program icon click** Double-click on the Windows desktop icon representing the program (in some Windows versions, only a single click may be needed).
- **Start Menu selection** Press the **Windows key** (or click on the onscreen Start button) and select the **Programs** or **All Programs** choice. From the menu list that appears, select the entry for the program. If the program is listed on a submenu, first select the submenu and then select the program.

Experienced Microsoft Windows users may know other ways to open a program. You can, of course, use these other ways if you prefer.

Program windows

Microsoft Windows is so called because, every time you open a program, a rectangular frame or ‘window’ opens on the screen. Inside this window, you interact with the program; many interactions lead to the opening of additional windows in which you make entries and selections.

The opening window of most programs contains these common elements:

- a **title bar** at the top of the window, identifying the program and any file in use
- **Resize buttons**, on the right side of the title bar area, that affect the displayed size of the window
- **Minimise buttons** that hide a program without closing it
- a **Close Program button** on the right margin of the title bar that allows you to end your use of a program quickly.

Some programs also contain:

- a **menu bar**, a horizontal list of words below the title bar that, when clicked, give pull-down lists of command choices
- one or more **toolbars** that contain buttons that are command shortcuts.

In the Office 2013 or multi-user Office 365 programs there is:

- a **quick access toolbar** at the top, which has buttons such as Undo and Redo
- **tabs** that display the names of key command areas
- **tab groups** that display collections of commands in related areas.

The area of tabs and related tab groups is known as the **Ribbon**. The first tab, File, takes you to the Backstage view, which provides access to certain commonly used commands such as opening, saving, sharing and printing.

Figure A1.1 shows these common elements for the opening window in Microsoft Excel 2013. Figure A1.2

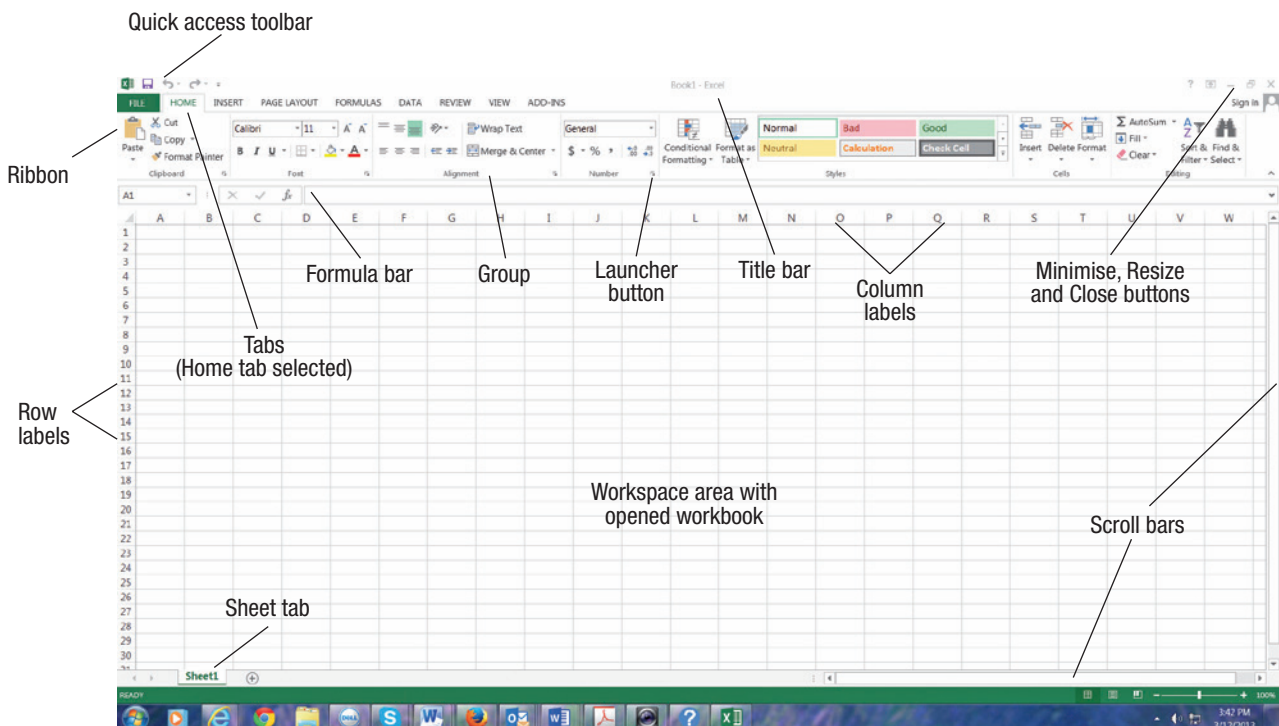


Figure A1.1 The Excel 2013 window

shows the opening window in Excel for Mac 2011. Note that the windows for this program have a worksheet area composed of rows and columns that you use for data entry. Figure A1.3 shows the Backstage view in Excel 2013 with a document about to be printed.

Dialog boxes

Many of the entries and selections you make as you use a program trigger the display of additional windows, known as dialog boxes. At the bottom corner of most tab groups a

small arrow will open a dialog box that provides additional choices. Many of the elements below are commonly found in dialog boxes, and some can be found in the Backstage view as shown in Figure A1.3.

- **Question mark help button** Clicking on this button allows you then to click on an element of the dialog box to display a help message about that element.
- **Drop-down list box** Displays a list of choices when you click on the drop-down button that appears at the right edge of the box.

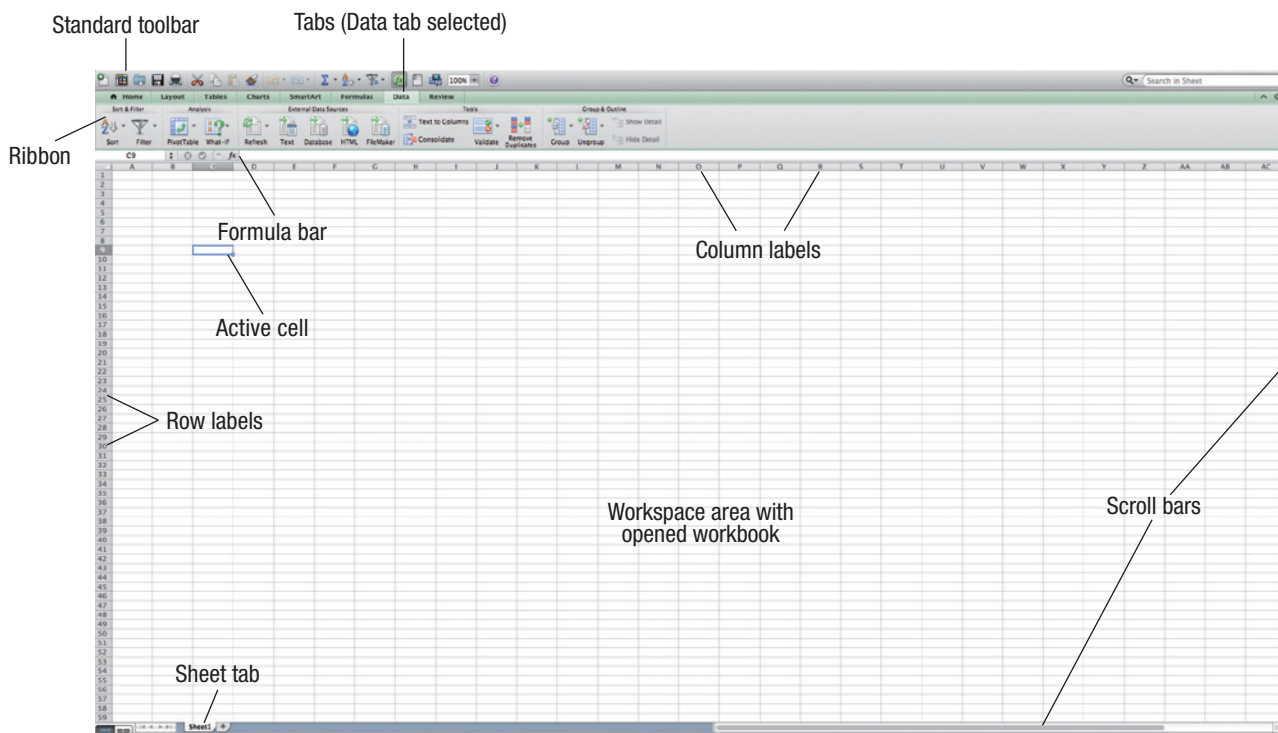


Figure A1.2 The Excel for Mac 2011 window

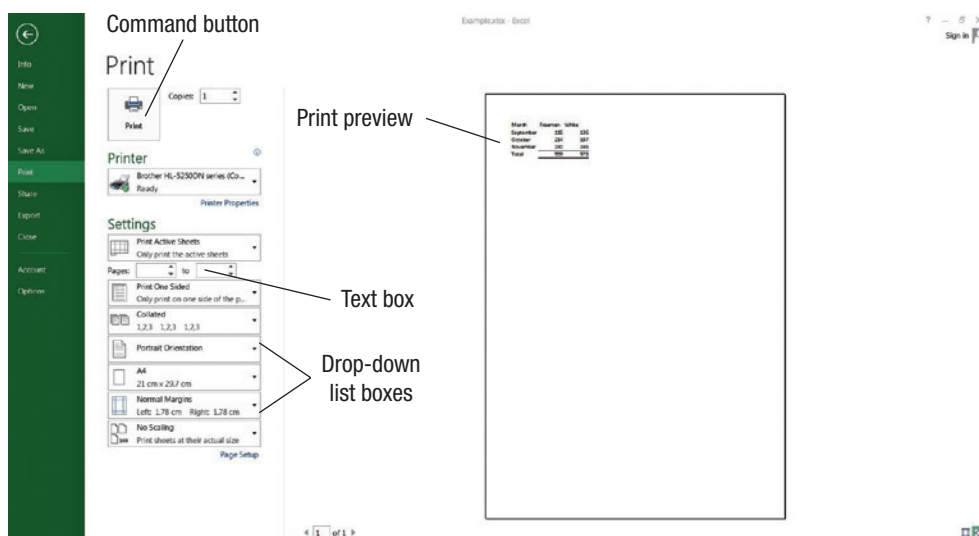


Figure A1.3 Microsoft Excel 2013 Backstage view with print selected

- **List box** Displays a list of choices. Sometimes includes **scroll buttons** or a **slider** if the list of choices is larger than the box can display.
- **Text box** Provides a space into which you can type an entry. These boxes are sometimes combined with either a drop-down list or *up* and *down* arrow buttons (seen in the Pages(s): ... to ... box in Figure A1.3), providing alternative ways of specifying an entry.
- **Check box** Provides a set of choices; you can make zero, one or more than one choice (compare with option buttons).
- **Icons** These allow you direct access to other places in the Windows system where files may be stored.
- **Option buttons** A set of mutually exclusive choices in which only one choice can be selected at a time.
- **Command buttons** Cause the program to take some action that usually closes the current dialog box and triggers the display of an additional dialog box. An **OK button** causes the program to take an action using the current values and settings of the dialog box. A **Cancel button** closes a dialog box and cancels the operation associated with that box.

How using Windows on a Mac differs

If you are running Windows on a Mac computer some of the information given above will not apply. Here are substitutes:

- You can open a program such as Excel by clicking on its symbol, the green X, on the **dock**.
- The **menu bar** appears at the top of the screen and contains commonly used commands such as File, Edit and View, which have drop-down menus.
- The **standard toolbar** shows the name of the open file and commonly used buttons such as **Save, Cut, Copy, Paste** and **Redo**.
- It also has three small circular red, yellow and green buttons above that allow you to **Close, Minimize** or **Maximise** a file.
- It may be necessary when using a program such as Excel to press Esc in order to see the toolbar and the three buttons.
- Alternatively, if you hover the cursor at the top of the screen a temporary toolbar will appear.
- Shortcut keys for copying are Command+c and for pasting Command+v. They can also be written as ⌘+c and ⌘+v.

Introduction to using Microsoft Excel on a PC

Microsoft Excel is the electronic worksheet program of Microsoft Office. Although not a specialised statistical program, Excel contains basic statistical functions and includes Data Analysis Toolpak procedures that you can use to perform selected advanced statistical methods. To use the Data Analysis Toolpak you must select it as an

Excel add-in. You can also install the PHStat2 add-in (available for separate purchase) to extend and enhance the Data Analysis Toolpak that Microsoft Excel contains. (You do not need to use PHStat2 in order to use Microsoft Excel with this text, although using PHStat2 will simplify using Excel for statistical analysis.)

In Microsoft Excel, you create or open and save files that are called **workbooks**. Workbooks are collections of worksheets and related items, such as charts, that contain the original data as well as the calculations and results associated with one or more analyses. Because of its widespread distribution, Microsoft Excel is a convenient program to use, but some statisticians have expressed concern about its lack of fully reliable and accurate results for some statistical procedures. Although Microsoft has recently improved many statistical functions, especially from Excel 2010 onwards, you should be somewhat cautious about using Microsoft Excel to perform analyses on data other than the data used in this text. (If you plan to install PHStat2, make sure you first read Appendix F and any PHStat2 read-me file.)

Opening and saving Microsoft Excel workbooks

Once you open the Excel program a new workbook will be displayed where you can begin entering data in rows and columns. If you wish to use a workbook created previously you will need to use the following commands.

If you are using Microsoft Excel 2013, select **File → Open**.

In the Backstage view you will be given a choice of selecting from Recent Workbooks, SkyDrive or the Computer. You can browse, select the file to be opened and then click on the **OK** button. If you cannot find your file, you may need to do one or more of the following:

- Use the scroll bars or the slider, if present, to scroll through the entire list of files.
- Select the correct folder from the drop-down list at the left-hand side of the dialog box.
- Change the **Files of type** value from the drop-down list at the bottom of the dialog box. Select **Text Files** from the list to see any text files; to list every file in the folder, select **All Files**.

In Excel 2013 select **File → Save As**, and in the Backstage view choose the location. In the dialog box enter (or edit) the name of the file in the **File name** box and click on the **OK** button. If applicable, you can also do the following:

- Change to another folder by selecting that folder from the **Save in** drop-down list.
- Change the **Save as type** value to something other than the default choice, **Microsoft Excel Workbook**. **Text (Tab delimited)** or **CSV (Comma delimited)** are two file types sometimes used to share Excel data with other programs.

After saving your work, you should consider saving your file a second time, using a different name, to create a backup copy of your work. Read-only files cannot be saved to their original folders unless the name is changed.

Microsoft Excel entries and references

In Microsoft Excel you enter data into worksheets that are organised as lettered columns and numbered rows. Typically, you enter the data for each variable in a separate column, using the row 1 cell for a variable label and each subsequent row for a single observation. You should follow the good practice of entering only one set of data per worksheet.

To refer to a specific entry, or cell, you use a *Sheetname!ColumnRow* notation. For example, Data!A2 refers to the cell in column A and row 2 in the Data worksheet. To refer to a specific group or **range** of cells, you use a *Sheetname!Upperleftcell:Lowerrightcell* notation. For example, Data!A2:B11 refers to the 20 cells that are in rows 2 to 11 in columns A and B of the Data worksheet. An absolute address for the cell A6 is shown as \$A\$6. Even if a formula using this address is copied to another row or column it will still refer to this cell. However, if the formula is written with the relative address A6, moving the formula will change the reference cell. Both absolute and relative addresses may be necessary in one sheet depending on the operations intended. Also note that \$A6 freezes the column but not the row and A\$6 freezes the row but allows the column to change.

Each Microsoft Excel worksheet has its own name. Automatically, Microsoft Excel names worksheets in the form of **Sheet1**, **Sheet2** and so on. You should rename your worksheets, giving them more self-descriptive names, by double-clicking on the sheet tabs that appear at the bottom of each sheet, typing a new name and pressing the Enter key.

Using formulas in Excel worksheets

Formulas are worksheet cell entries that perform a calculation or some other task. You enter formulas by typing the equals sign symbol (=) followed by some combination of mathematical or other data-processing operations.

For simple formulas, you use the symbols +, −, *, / and ^ for the operations addition, subtraction, multiplication, division and exponentiation (a number raised to a power), respectively. For example, the formula

$$=\text{Data!B2} + \text{Data!B3} + \text{Data!B4} + \text{Data!B5}$$

adds the contents of the cells B2, B3, B4 and B5 of the Data worksheet and displays the sum as the value in the cell containing the formula. You can also use Microsoft Excel *functions* in formulas to simplify formulas. To find lists of the functions that can be selected in Excel, click on the f_x Function Wizard symbol on the Formula bar. For example, the formula =SUM(Data!B2:B5), using the

Excel SUM() function, is a shorter equivalent to the formula above.

You can also use cell or cell range references that do not contain the *Sheetname!* part, such as B2 or B2:B5. Such references *always* refer to the worksheet in which the formula has been entered.

Formulas allow you to create generalised solutions and give Excel its distinctive ability to recalculate results automatically when you change the values of the supporting data. Typically, when you use a worksheet, you see only the results of any formulas entered, not the formulas themselves. However, for your reference, many illustrations of Microsoft Excel worksheets in this text also show the underlying formulas adjacent to the results they produce. When using Excel 2013 select **Formulas** → **Formula Auditing** → **Show Formulas** to see onscreen the formulas themselves and not their results. To restore the original view, click on **Show Formulas** again.

Creating charts

The method of creating charts varies according to the version of Excel you are using.

- Excel 2013** A new feature in Excel 2013 allows you to create charts easily using the **Quick Analysis** tool. Simply highlight an area of the spreadsheet containing some data you wish to graph by clicking on the top left-hand cell, then dragging the mouse. The range may contain labels. Click on the small box that appears in the bottom right-hand corner to open **Quick Analysis**. Select **Charts**, then, by hovering the mouse over the different chart types, you can see previews of recommended charts for the selected data. You can also choose **More Charts**, which will open a dialog box with a more extensive range of options. Once a chart is selected there are several ways you can modify it by clicking on the three icons that appear on its right-hand side. These are **Chart Elements** (+), **Chart Styles** (paintbrush) and **Chart Filters** (filter). You will also now see that multiple design options are shown on the ribbon and that options to change colours or chart type are shown there. By right-clicking on the background area of the chart you can also activate a drop-down menu. If you choose **Format Chart Area** a menu will open on the right-hand side of the spreadsheet that allows you to change the format of the chart and text in many ways. If instead you choose **Move Chart** you can choose a new location on another sheet. To reposition the chart on the existing sheet, simply click on it and drag.
- Excel 2010** Highlight the area of the spreadsheet with your data as described above. If you wish to select areas that are not adjacent, hold down the Ctrl key while selecting. The area selected must be rectangular. Click on the **Insert** tab, then from the Charts area click on the

type of chart required and select a particular format from the drop-down gallery. The chart can be enhanced by clicking on it, choosing **Layout** and then selecting formatting icons such as **Chart Title**. Other changes may be made by right-clicking on the chart to access **Select Data** and **Format Plot Area** menus or to **Move Chart**.

Figure A1.4 shows an example of a chart created in Excel 2013 with the Format Chart Area panel open.

Printing workbooks

Before printing you may select a print area if you do not want the whole sheet printed. To print Excel 2013 worksheets, select **File** → **Print**. A print preview is automatically created, as can be seen in Figure A1.3. Various print settings are available in the drop-down list boxes, and clicking on Page Setup will give access to more choices. When you are satisfied with the settings and look of the preview click on the Print button.

Note that if you want only a part of the worksheet to be printed it is easier to set this using **Page Layout** tab then **Page setup** → **Print area**.

Page Setup allows you to customise printing to change the print orientation, add gridlines and so on before printing. Once you are satisfied with the results, click on the **Print** button in the print preview window, then **OK** in the **Print** dialog box.

The Print Backstage view (see Figure A1.3) contains settings to select the printer to be used, what parts of the workbook to print (the active worksheet is the default) and the number of copies to produce (1 is the default). If you

need to change these settings, change them before clicking on the **OK** button.

After printing, you should verify the contents of your printout. Most printing failures will trigger the display of an error message that you can use to work out the source of the failure.

How using Excel for Mac 2011 differs

Excel for Mac 2011 does not come with add-ins for Analysis Toolpack. However, it is possible to download software made by third-party companies to perform some of the same statistical analysis tasks. The free program StatPlus®:mac LE, for instance, will allow you to run a regression, calculate descriptive statistics and run analysis of variance tests. Further capability is available in the Pro edition at a cost.

You can open a new workbook by using the Excel Workbooks Gallery → Templates → All → Excel Workbook. Alternatively click on File → New Workbook.

The easiest way to save a new workbook is to use Command+s keys on the Mac keyboard (i.e. ⌘+s). A **Save As** dialog box will allow you to choose a file name, a location for the file and the file format. You can also choose File → Save to begin this process.

To create a chart in Excel for Mac 2011 you need to first highlight the range that contains the required data, then choose the **Charts** tab. You can then choose from various chart types in the **Insert Chart** group. You are able to view various subtypes and check which one displays your data best. You will find that extra tabs **Chart Layout** and **Chart Format** open on the ribbon to allow more formatting

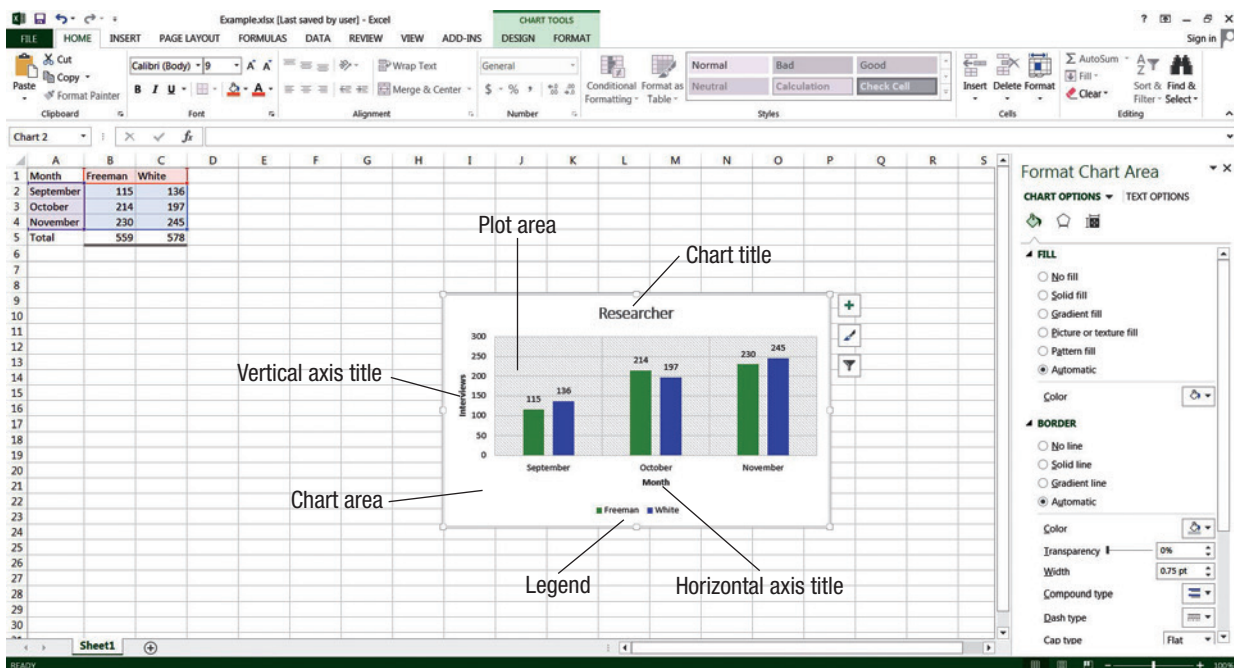


Figure A1.4 Microsoft Excel 2013 showing a chart and the formatting panel

options. For instance, you can add axis titles by following the path **Chart Layout** → **Labels** → **Axis Titles**.

Shortcut keys for printing a file are Command+p (i.e. ⌘+p). This will bring up a dialog box, a print preview and

various print options. See Figure A1.5 for a screenshot of a worksheet containing data, a chart constructed from the data, and the printing dialog box with print preview and printing options.

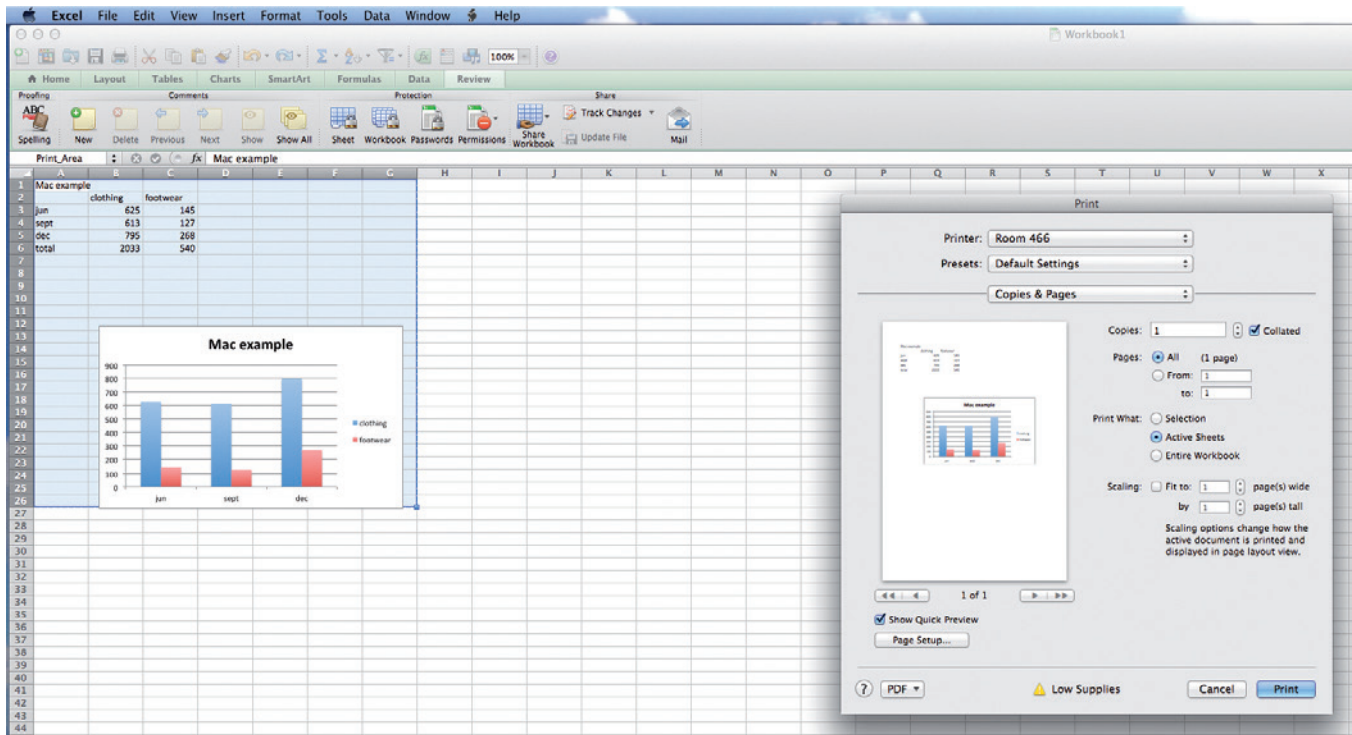


Figure A1.5 Printing dialog box in Excel for Mac 2011

Microsoft® product screen shots are reprinted with permission from Microsoft Corporation.

2

CHAPTER

Presenting data in tables and charts

Festival expenditure

A council is investigating the contribution to the local economy of visitors to an annual three-day music festival. Kai, a researcher employed by the council, has collected data from a random sample of non-local festival attendees aged 18 years and over. This data includes total amount spent, excluding festival tickets, in the region during the festival and whether the festival attendee has travelled from within the state (intrastate), from another state (interstate) or from another country (international) to attend the festival.

The data is stored in the < **FESTIVAL** > file.

Kai is interested in answering the following questions:

- What is the typical amount spent during the festival by intrastate, interstate and international visitors?
- How does the amount spent vary between visitors and between intrastate, interstate and international visitors?
- Is there a difference in the amount spent between intrastate, interstate and international visitors?

© Zoonar/Thomas Willer/age fotostock



learning objectives

After studying this chapter you should be able to:

- 1 describe the distribution of a single categorical variable using tables and charts
- 2 describe the distribution of a single numerical variable using tables and graphs
- 3 describe the relationship between two categorical variables using contingency tables
- 4 describe the relationship between two numerical variables using scatter diagrams and time-series plots
- 5 correctly present data in graphs

Kai needs to organise the data into usable forms. One way of doing this is to use tables or charts. This chapter helps you to select and construct appropriate tables and charts. We can also use numerical measures to determine certain characteristics of the data, such as their centre and spread. These numerical descriptive measures are covered in the next chapter.

From Chapter 1 we know that data can be either categorical or numerical.

2.1 TABLES AND CHARTS FOR CATEGORICAL DATA

The expenditure data in the < **FESTIVAL** > file are examples of raw data – that is, data presented just as they were collected. Raw data give very little information, but by using summary tables and charts we can condense and present the data in a meaningful way. For categorical data, you first divide the data into categories and then present the frequency or percentage in each category in a table or chart.

Summary Tables

A **summary table** gives the frequency, proportion or percentage of the data in each category, which allows you to see differences between the categories. A summary table lists the categories in one column and the frequency, percentage or proportion in a separate column or columns.

Table 2.1 illustrates a summary table based on a recent survey that asked why people shopped for groceries online. In Table 2.1 the most common reason for grocery shopping online was convenience, followed by competitive prices and quality products. Very few respondents shopped for groceries online because of a comfortable environment or well-displayed products.

Reason	Percentage
Comfortable environment	8
Competitive prices	20
Convenience	28
Customer service	13
Products well displayed	3
Quality products	18
Variety/range of products	10
	100

LEARNING OBJECTIVE

1

Describe the distribution of a single categorical variable using tables and charts

summary table

Summarises categorical or numerical data; gives the frequency, proportion or percentage of data values in each category or class.

Table 2.1

Reasons for grocery shopping online

EXAMPLE 2.1

SUMMARY TABLES FOR LOCATION AND TYPE OF PROPERTIES

In other research Kai is exploring the property market in the council area. Data from 100 recent property sales is stored in <PROPERTY>. These properties are classified according to location, either in town or rural, and also by type, either a house or a unit. Construct summary tables for the properties categorised by location and type.

SOLUTION

Table 2.2A

A frequency and percentage summary table for the location of 100 recent property sales

Location	Number (frequency) of properties	Percentage of properties
Rural	34	34.0
Town	66	66.0
Total	100	100.0

From Table 2.2A we can see that there are approximately twice as many urban properties sold as rural properties.

Table 2.2B

A frequency and percentage summary table for type of 100 recent property sales

Type	Number of properties	Percentage of properties
House	82	82.0
Unit	18	18.0
Total	100	100.0

From Table 2.2B we can see that there are relatively few units sold.

Bar Charts

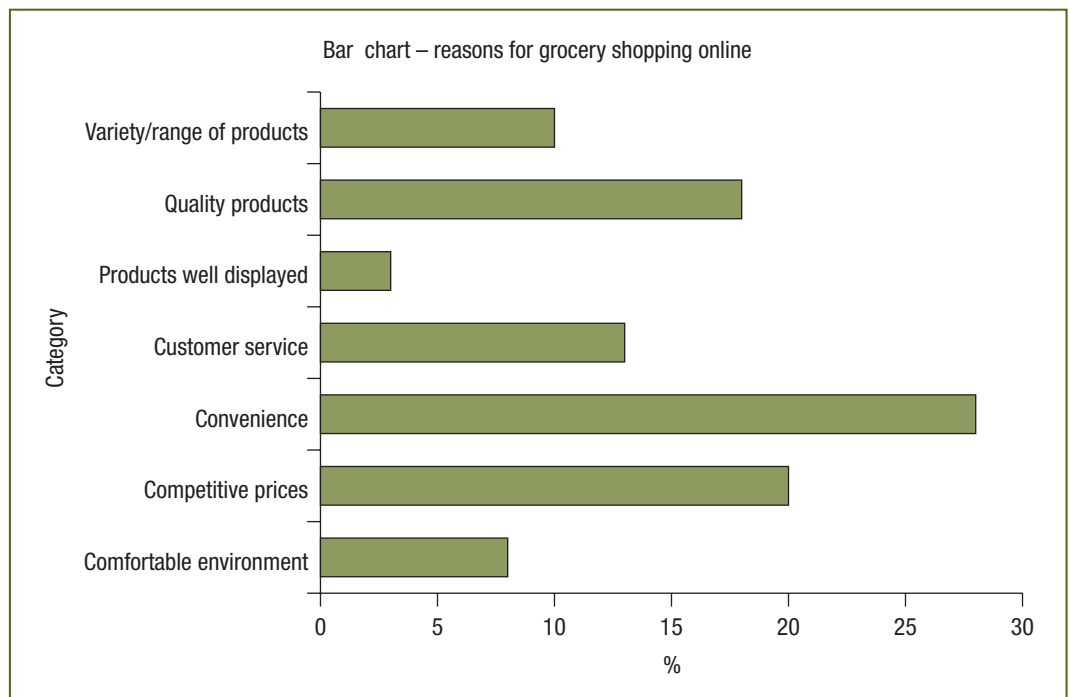
bar chart

Graphical representation of a summary table for categorical data; the length of each bar represents the proportion, frequency or percentage of data values in a category.

Each category in a **bar chart** is represented by a bar, the length of which indicates the proportion, frequency or percentage of values falling into that category. Figure 2.1 displays a bar chart of the reasons for grocery shopping online, presented in Table 2.1. Bar charts allow you to compare percentages, frequencies or proportions in the different categories. In Figure 2.1 the most common reason for shopping online is convenience, followed by competitive prices. Very few respondents shopped for groceries online because of a comfortable environment or well-displayed products.

Figure 2.1

Microsoft Excel bar chart of the reasons for grocery shopping online

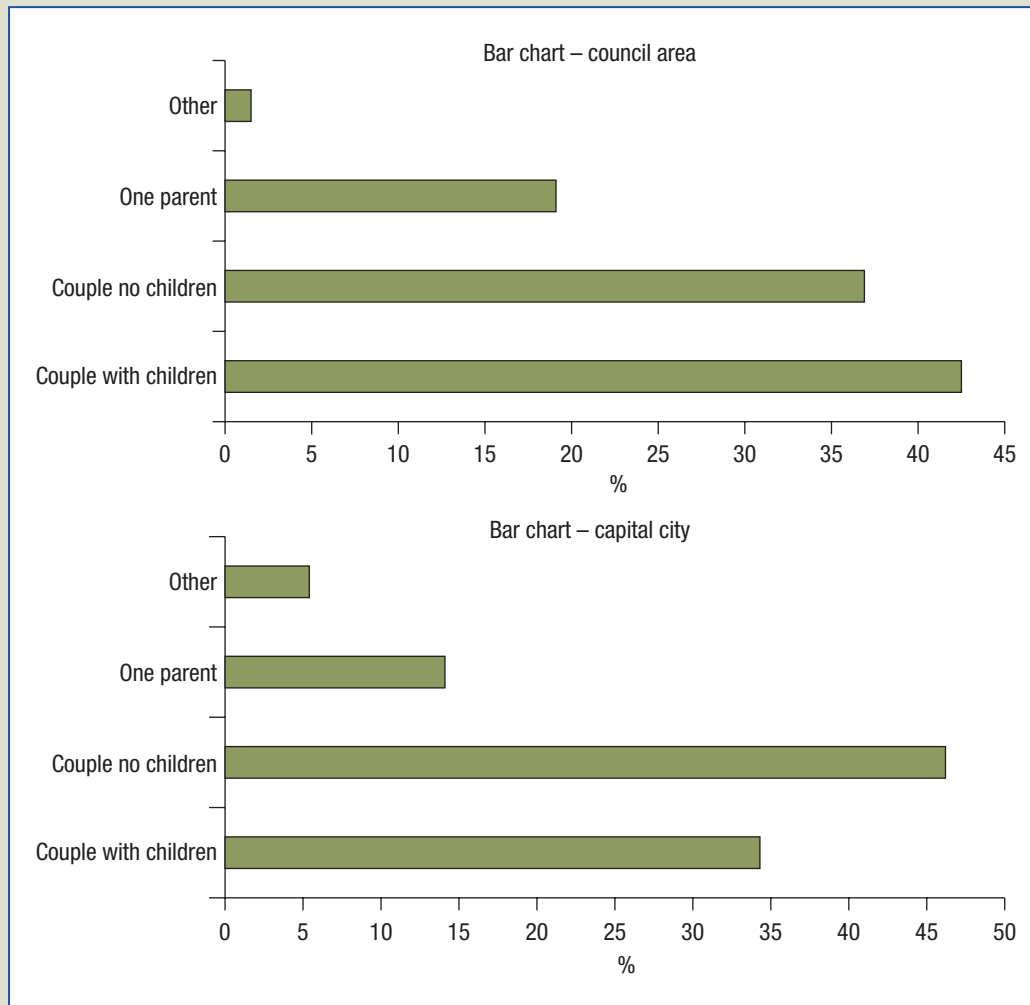


BAR CHART FOR FAMILY TYPE

The council is also interested in demographic differences between the council area and the capital city. Demographic information has been collected and is stored in < DEMOGRAPHIC_INFORMATION >

Use the summary tables for family type to construct and interpret bar charts for the council area and the capital city.

SOLUTION



We can see that the majority of families are couples with or without children in both areas, with a significant number of one-parent families. However, the capital city has approximately 10% more couples without children and 5% fewer one-parent families.

EXAMPLE 2.2

Figure 2.2

Microsoft Excel bar chart for family type

Pie Charts

A **pie chart** is a circle, used to represent the total, which is divided into slices, each representing a category. The area of each slice represents the proportion or the percentage share of the corresponding category. In Table 2.1, for example, 28% of the respondents said that convenience was the main reason for grocery shopping online. Thus, in constructing the pie chart, the 360° that makes up a circle is multiplied by 0.28, resulting in a slice of the pie that takes up 100.8° of the 360° of the circle (Figure 2.3). A pie chart allows you to see the portion of the entire pie that falls into each category. In Figure 2.3, convenience takes 28% of the pie and products well displayed takes only 3%.

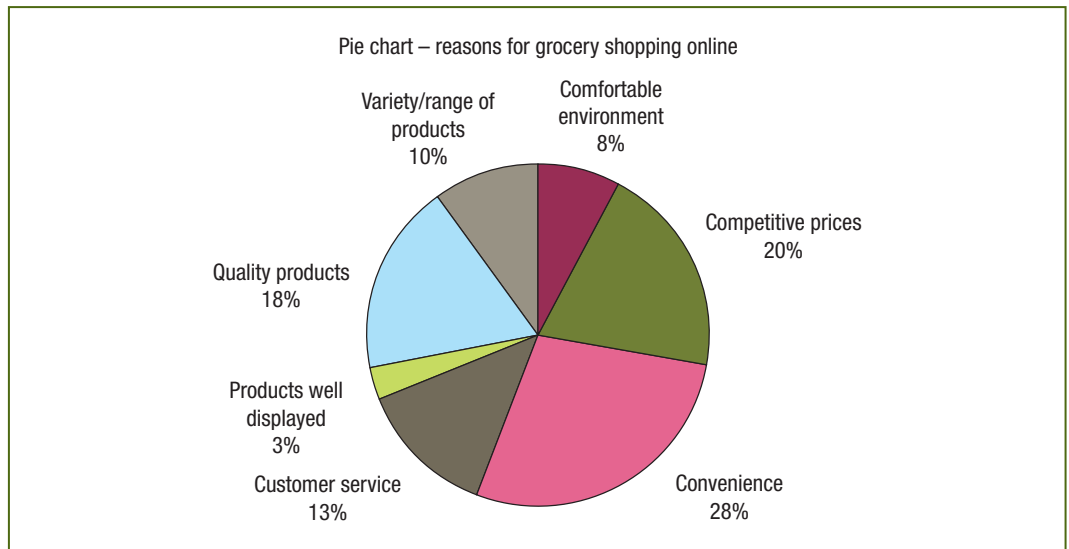
pie chart

Graphical representation of a summary table for categorical data, each category represented by a slice of a circle of which the area represents the proportion or percentage share of the category relative to the total of all categories.

What type of chart should you use? The selection of a chart depends on your intention. If a comparison of categories is most important, use a bar chart. If observing the portion of the whole that lies in a particular category is most important, use a pie chart.

Figure 2.3

Microsoft Excel pie chart of the reasons for grocery shopping online



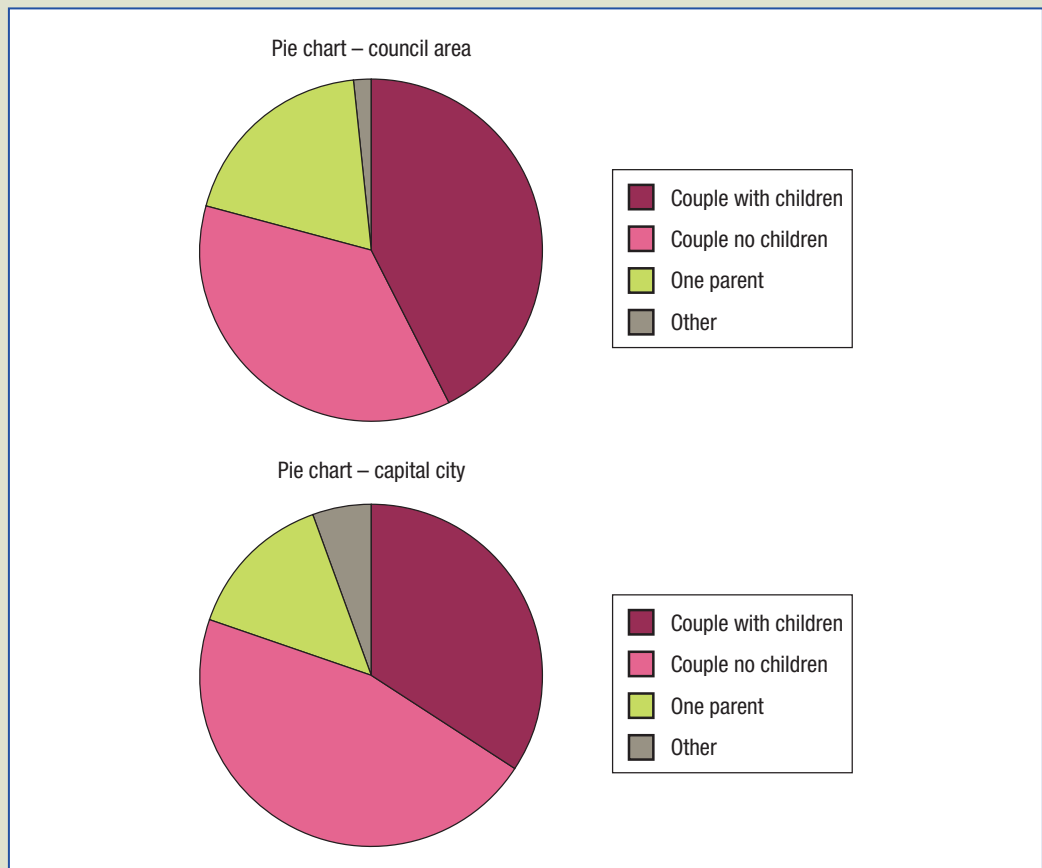
EXAMPLE 2.3

PIE CHART FOR FAMILY TYPE

The council and capital city demographics are presented as summary tables. Use the summary tables given for family type to construct and interpret pie charts for the capital city and the council area.

Figure 2.4

Microsoft Excel pie chart for family type



We can see that most families are couples with or without children in both areas, with a significant number of one-parent families. However, the city has a higher proportion of couples without children.

Problems for Section 2.1

LEARNING THE BASICS

- 2.1** A categorical variable has three categories with the following frequency of occurrence:

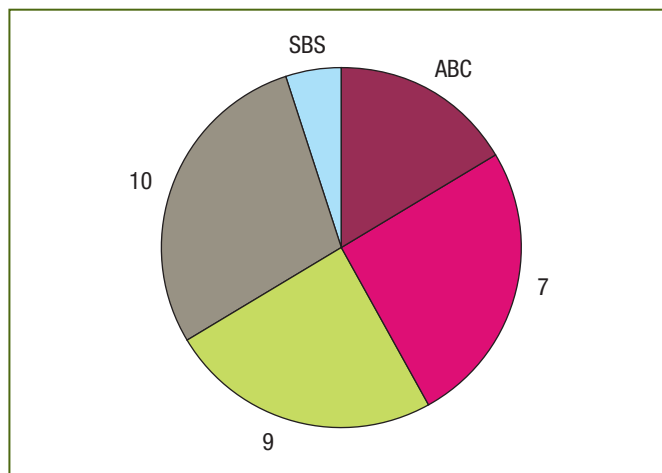
Category	Frequency
A	13
B	28
C	9

- Calculate the percentage of values in each category.
 - Construct a bar chart.
 - Construct a pie chart.
- 2.2** A categorical variable has four categories with the following percentages of occurrence:

Category	Percentage	Category	Percentage
A	12	C	35
B	29	D	24

- Construct a bar chart.
- Construct a pie chart.

2.3



The pie chart above was constructed from the results of a survey of 2,000 viewers to determine which TV channels they watch for news. By measuring the angle of each one using a protractor, or estimating by eye, calculate the percentage of viewers watching:

- ABC
- Channel 7
- Channel 9

- Channel 10
- SBS

APPLYING THE CONCEPTS

You can solve problems 2.4 to 2.7 manually or by using Microsoft Excel.

- 2.4** The following table gives the top 10 websites ranked by number of unique visitors for December 2010.

Domain	Total unique visitors (millions)
google.com	149.98
facebook.com	142.18
yahoo.com	136.95
youtube.com	117.21
amazon.com	92.53
bing.com	79.87
wikipedia.org	79.80
live.com	74.95
msn.com	74.61
ebay.com	69.66

Data obtained from Compete Inc., Clickstream panel <<http://blog.compete.com/2011/02/07/facebook-inches-closer-to-the-top-in-compete-ranking-of-top-50-web-sites-for-december-2010>> accessed January 2011

- Construct a bar chart and a pie chart.
 - Which graphical method do you think best portrays these data?
 - What conclusions can you reach concerning the number of unique visitors?
- 2.5** Pat of Pat's Cars asked 200 customers their colour preference when purchasing a new car. The following summary table gives the results.

Colour	Frequency
White	56
Blue	31
Red	29
Brown	17
Grey	19
Silver	15
Green	15
Black	13
Other	5

- Construct a pie chart and a bar chart.
 - What colours of cars should Pat have on show?
- 2.6** The following table gives the labour force status of the Australian civilian population aged 15 years and over in November 2013.

6202.0 - Labour Force, Australia, Nov 2013	
Labour force status (aged 15 years & over)	Total '000
Employed full-time	8,112.5
Employed part-time	3,522.7
Unemployed looking for full-time work	492.1
Unemployed not looking for full-time work	173.1
Not in labour force	6,788.2
Civilian population 15 aged years and over	19,088.6

Data obtained from Australian Bureau of Statistics, Labour Force, Australia, Nov 2013, Cat. No. 6202.0 <www.abs.gov.au/ausstats/abs@.nsf/mf/6202.0> accessed 22 December 2013

- a. For the data construct a bar chart and a pie chart.
 - b. Which graphical method do you think best portrays these data?
 - c. What conclusions can you draw about participation rate – that is, the percentage of the population in the labour force?
- 2.7 Use the summary table for country of birth in the < DEMOGRAPHIC_INFORMATION > file to construct pie and bar charts.

2.2 ORGANISING NUMERICAL DATA

LEARNING OBJECTIVE

2

Describe the distribution of a single numerical variable using tables and graphs

When you have a large amount of numerical raw data, a useful first step is to present the data as either an ordered array or a stem-and-leaf display.

Suppose you undertake a study to compare the cost of a main meal at similar restaurants in a city and in the suburbs. Table 2.3 gives the raw data for 50 city restaurants and 50 suburban restaurants; these data are stored in the < RESTAURANT > file. From the raw data it is difficult to draw any conclusions about the price of city and suburban restaurant meals.

Table 2.3 Price per main meal at 50 city restaurants and 50 suburban restaurants

City									
50	38	43	56	51	36	25	33	41	44
34	39	49	37	40	50	50	35	22	45
44	38	14	44	51	27	44	39	50	35
31	34	48	48	30	42	26	35	32	63
36	38	53	23	39	45	37	31	39	53
Suburban									
37	37	29	38	37	38	39	29	36	38
44	27	24	34	44	23	30	32	25	29
43	31	26	34	23	41	32	30	28	33
26	51	26	48	39	55	24	38	31	30
51	30	27	38	26	28	33	38	32	25

Ordered Arrays

A more meaningful display is obtained by sorting the raw data in order of magnitude – that is, from smallest to largest. This is called an **ordered array**. Table 2.4 presents the data in Table 2.3 as an ordered array. From Table 2.4 you can see that the price of a main meal at city restaurants is between \$14 and \$63, and the price of a main meal at suburban restaurants is between \$23 and \$55.

Stem-and-Leaf Displays

A **stem-and-leaf display** is a quick and easy way of displaying numerical data visually. The data are divided into groups (called stems) such that the values within each group (the leaves) branch out to the right on each row. The resulting display allows you to see how the data are distributed and also where they are concentrated.

ordered array

Numerical data sorted by order of magnitude.

stem-and-leaf display

Graphical representation of numerical data; partitions each data value into a stem portion and a leaf portion.

.....
Table 2.4 Ordered array of price per main meal at 50 city restaurants and 50 suburban restaurants

City									
14	22	23	25	26	27	30	31	31	32
33	34	34	35	35	35	36	36	37	37
38	38	38	39	39	39	39	40	41	42
43	44	44	44	44	45	45	48	48	49
50	50	50	50	51	51	53	53	56	63
Suburban									
23	23	24	24	25	25	26	26	26	26
27	27	28	28	29	29	29	30	30	30
30	31	31	32	32	32	33	33	34	34
36	37	37	37	38	38	38	38	38	38
39	39	41	43	44	44	48	51	51	55

To see how a stem-and-leaf display is constructed, suppose that 20 students eat lunch at a fast-food restaurant, spending the following amounts:

\$6.35 \$4.75 \$4.30 \$5.40 \$4.85 \$6.60 \$5.55 \$4.90 \$6.85 \$7.50
 \$8.45 \$6.05 \$9.90 \$5.75 \$6.80 \$4.30 \$5.45 \$7.20 \$7.80 \$10.65

To construct a stem-and-leaf display for these data, use the \$ amount as the stem and round the cents to the nearest 10 cents for the leaves. Now list the stem values (\$) in order of size to the left of a vertical divider (|) and then record the leaves (10 cents) for each stem in rows to the right. The 'unordered' stem-and-leaf display for the amount spent on lunch by the 20 students is:

stem unit: \$	leaf unit: 10 cents
4	8 3 9 9 3
5	4 6 8 5
6	4 6 9 1 8
7	5 2 8
8	5
9	9
10	7

The first value of \$6.35 is rounded to 6.4. Its stem (row) is 6 and its leaf is 4. The second value of \$4.75 is rounded to 4.8. Its stem (row) is 4 and its leaf is 8. Then, ordering each leaf, we obtain the following ordered stem-and-leaf display for the amount spent on lunch by the 20 students:

stem unit: \$	leaf unit: 10 cents
4	3 3 8 9 9
5	4 5 6 8
6	1 4 6 8 9
7	2 5 8
8	5
9	9
10	7

STEM-AND-LEAF DISPLAY FOR FESTIVAL EXPENDITURE – INTERSTATE VISITORS

Kai is interested in the amount spent during the festival by interstate visitors. < **FESTIVAL** >

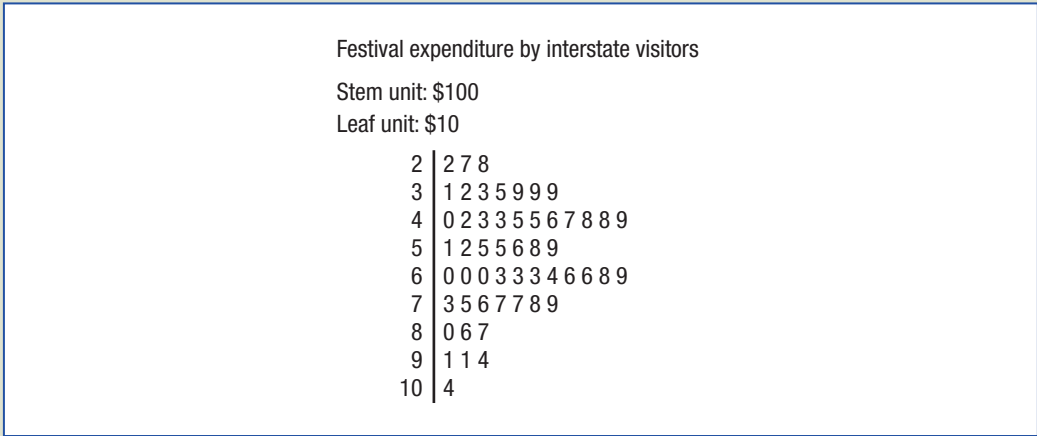
Construct and interpret a stem-and-leaf display for these data.

EXAMPLE 2.4

SOLUTION

Figure 2.5

PhStat2 stem-and-leaf display for festival expenditure by interstate visitors



From Figure 2.5 Kai can conclude that during the festival:

- interstate visitors spend between \$220 and \$1,040
- most interstate visitors spend between \$300 and \$800
- interstate visitors rarely spend less than \$300 or more than \$800.

Problems for Section 2.2

LEARNING THE BASICS

2.8 Form an ordered array given the following data from a sample of $n = 7$ mid-semester exam scores in accounting:

68 94 63 75 71 88 64

2.9 Form a stem-and-leaf display given the following data from a sample of $n = 7$ mid-semester exam scores in finance:

80 54 69 98 93 53 74

2.10 Form an ordered array given the following stem-and-leaf display from a sample of $n = 7$ mid-semester exam scores in information systems:

stem unit: 10	leaf unit: 1
5	0
6	
7	4 4 6
8	1 9
9	2

APPLYING THE CONCEPTS

2.11 Data were collected on the monthly expenses submitted by 35 employees in a firm’s sales team. The data are summarised in the following stem-and-leaf display:

stem unit: \$100	leaf unit: \$10
1	1 2 4 8 9
2	0 0 1 3 9 9 9 9 9 9
3	0 1 1 2 4 4 4 5 8 9 9
4	1 1 5 5 6
5	0 1 5 6

- Place the data into an ordered array.
- Which of the two displays provides the most information? Discuss.
- In what range are most monthly expense claims?
- Is there a concentration of expense claims near the centre of the distribution?

2.12 The following data represent the bounced cheque fee in dollars for a sample of 22 transaction accounts.

< BOUNCED_CHEQUE >

20 40 40 38 35 35 45 50 45 40 35
45 20 38 45 45 15 35 40 35 45 40

- Display the data as an ordered array.
- Construct a stem-and-leaf display for the data.
- Which of the two displays provides the most information? Discuss.
- Around what value, if any, are the bounced cheque fees concentrated? Explain.

2.13 The following data represent ATM withdrawal fees for withdrawals made above the free monthly allowance for a sample of 26 transaction accounts. < **ATM_FEE** >

0.65 0.50 0.70 1.30 2.50 0.50 2.00 1.00 2.00 1.25 1.50 2.00 0.30
2.00 0.65 2.00 0.50 0.65 0.50 0.65 1.60 0.70 1.00 1.50 1.65 0.50

Data obtained from <moneymanager.smh.com.au>

- Display the data as an ordered array.
 - Construct a stem-and-leaf display for the data.
 - Which of the two displays provides the most information? Discuss.
 - Around what value, if any, are the withdrawal fees concentrated? Explain.
- 2.14** Low-fat foods are not necessarily low calorie, as many low-fat foods are high in sugar. The following data give calories per 250 ml cup of a random sample of brands of fresh cow's milk for sale in Australia. < **FRESH_MILK** >

Full cream milk

155 188 160 155 160 163 170 185 135 160 165 160 163

Low- or reduced-fat milk

120 133 133 125 118 113 140 110 128 115

No-fat or skim milk

133 90 90 98 88 85 115 108 88 90 90 98

Data obtained from Calorie King Australia <www.calorieking.com.au> accessed 22 December 2013

For each category of milk:

- Display the data in ordered arrays.
- Construct stem-and-leaf displays for the data.
- Which arrangement provides more information? Discuss.
- Compare the items in terms of calories. What conclusions can you make?

2.3 TABLES AND GRAPHS FOR NUMERICAL DATA

Ordered arrays and stem-and-leaf displays are of limited use when we have very large quantities of data or the data are highly variable. In these cases we use tables and graphs to condense and present the data visually. These tables and graphs include histograms, frequency, relative frequency, and cumulative distributions and polygons.

Frequency Distributions

A frequency distribution allows you to condense a set of data.

A **frequency distribution** is a summary table in which the data are arranged into numerically ordered classes or intervals.

To construct a frequency distribution, first select an appropriate *number* of classes and a suitable **class width**. The classes should be exhaustive and mutually exclusive, so that any one data value belongs to one and only one class. The number of classes chosen depends on the amount of data – a small number of classes for small amounts of data and a larger number of classes for larger amounts of data. In general, a frequency distribution should have at least five classes but no more than 15. If there are too few classes we lose too much information and if there are too many classes the data are not condensed enough.

Each class should be of equal width. To determine the required (approximate) width of the classes, divide the **range** (the highest value – the lowest value) of the data by the required number of classes.

DETERMINING THE WIDTH OF A CLASS

$$\text{Class width} = \frac{\text{range}}{\text{number of classes}} \quad (2.1)$$

LEARNING OBJECTIVE

2

Describe the distribution of a single numerical variable using tables and graphs

frequency distribution

Summary table for numerical data; gives the frequency of data values in each class.

class width

Distance between upper and lower boundaries of a class.

range

Distance measure of variation; difference between maximum and minimum data values.

The city restaurant data consist of a sample of 50 restaurants; with this sample size 10 is an appropriate number of classes. From the ordered array in Table 2.4, the range of the data is \$63 – \$14 = \$49. Using Equation 2.1, the approximate class width is:

$$\text{Class width} = \frac{49}{10} = 4.9$$

Choose a class width that simplifies the reading and interpretation of the distribution and resultant graphs. Therefore, instead of using a class width of \$4.90, choose a width of \$5.00.

class boundaries

Upper and lower values used to define classes for numerical data.

Construct the frequency distribution table by first establishing clearly defined **class boundaries** so that each data value belongs in one and only one class. The classes must be *mutually exclusive* and *exhaustive*. Whenever possible, choose class boundaries that simplify the reading and interpretation of the resultant tables or graphs. For the city restaurant data the price ranges from \$14 to \$63, so appropriate classes could be (1) from \$10 to less than \$15, (2) from \$15 to less than \$20, and so on, until we have included the highest data value, in this case \$63. The last and 11th class ranges from \$60 to less than \$65. The centre of each class, called the **class mid-point**, is halfway between the lower boundary and the upper boundary of the class. Thus, the class mid-point for the first class, from \$10 to under \$15, is \$12.50 $\left(\frac{10 + 15}{2}\right)$; the class mid-point for the second class, from \$15 to under \$20, is \$17.50, and so on. Table 2.5 gives a frequency distribution of the cost per meal for the 50 city and the 50 suburban restaurants.

class mid-point

Centre of a class; representative value of class.

A frequency distribution allows you to draw conclusions about the major characteristics of the data. For example, Table 2.5 shows that the price of main meals at city restaurants is concentrated between \$30 and \$55 compared with the price of main meals at suburban restaurants, which are clustered between \$25 and \$40.

For small data sets, one set of class boundaries may provide a different picture from another set. For example, for the restaurant price data, using a class width of 4.0 instead of 5.0 (as was used in Table 2.5) may cause shifts in the way in which the values are distributed among the classes.

You can also get shifts in data concentration when you choose different lower and upper class boundaries. Fortunately, as the sample size increases, alterations in the selection of class boundaries affect the concentration of data less and less.

Table 2.5 Frequency distribution of the price per main meal for 50 city restaurants and 50 suburban restaurants

Price of main meal (\$)	City frequency	Suburban frequency
\$10 but less than \$15	1	0
\$15 but less than \$20	0	0
\$20 but less than \$25	2	4
\$25 but less than \$30	3	13
\$30 but less than \$35	7	13
\$35 but less than \$40	14	12
\$40 but less than \$45	8	4
\$45 but less than \$50	5	1
\$50 but less than \$55	8	2
\$55 but less than \$60	1	1
\$60 but less than \$65	1	0
Total	50	50

EXAMPLE 2.5

FREQUENCY DISTRIBUTION FOR FESTIVAL EXPENDITURE – INTERSTATE VISITORS
 Kai is interested in the amount spent during the festival by interstate visitors. **< FESTIVAL >**
 Construct and interpret a frequency distribution for this data.