5TH EDITION

Basic Business Statistics

Concepts and applications

Berenson Levine Szabat
O'Brien Jayne Watson

Pearson

5TH EDITION

Basic Business Statistics

This page is intentionally left blank

5TH EDITION

Basic Business Statistics

Concepts and applications

Berenson Levine Szabat
O'Brien Jayne Watson

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.

Fifth adaptation edition published by Pearson Australia Group Pty Ltd, Copyright © 2019

Portfolio Manager: Rebecca Pedley
Development Editor: Anna Carter
Project Managers: Anubhuti Harsh and Keely Smith
Production Manager: Julie Ganner
Product Manager: Sachin Dua
Content Developer: Victoria Kerr
Rights and Permissions Team Leader: Lisa Woodland
Lead Editor/Copy Editor: Julie Ganner
Proofreader: Katy McDevitt
Indexer: Garry Cousins
Cover and internal design by Natalie Bowra
Cover photograph © kireewong foto/Shutterstock
Typeset by iEnergizer Aptara®, Ltd

Printed in Malaysia

ISBN 9781488617249

1 2 3 4 5 23 22 21 20 19

A catalogue record for this book is available from the National Library of Australia

Pearson Australia Group Pty Ltd ABN 40 004 245 943

brief contents

detailed contents

preface

This fifth Australasian and Pacific edition of *Basic Business Statistics: Concepts and Applications* continues to build on the strengths of the fourth 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 fourth 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, Chapter 19 on further non-parametric tests and two brand new chapters: Chapter 20 on business analytics and Chapter 21 on data analysis. This chapter will be especially useful to students who wish to understand how the concepts and techniques studied in this book all fit together. The Part 5 chapters can be found within the MyLab and student download page via our catalogue.

Chapter 21 (including Figure 21.1, which provides a summary of the contents of this book arranged by data-analysis task) is designed to provide guidance in choosing appropriate statistical techniques to data-analysis questions arising in business or elsewhere. Figure 21.1, and Chapter 21, should be referred to when working through the earlier chapters of this book. This should enable students to see connections between topics; that is, the big picture.

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.

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 are:

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.

Technical Editor

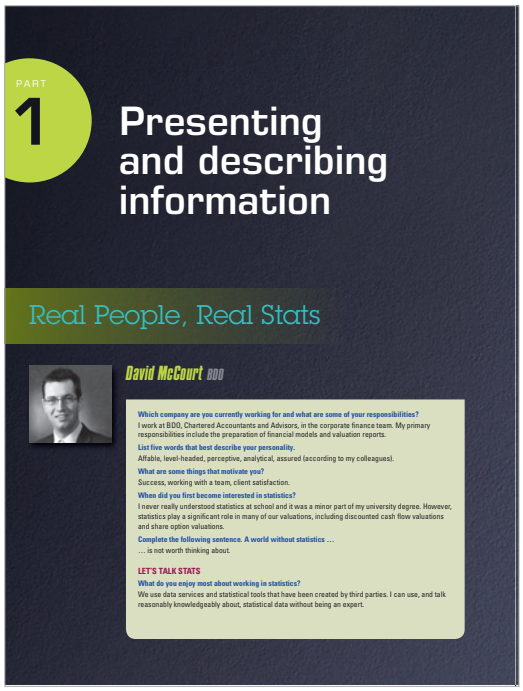
We would like to thank Martin Firth at UWA for carrying out a detailed technical edit of the text.

Reviewers

Ms Gerrie Roberts *Monash University*
Dr Sonika Singh *University of Technology Sydney*
Dr Erick Li *University of Sydney*
Dr Amir Arjomandi *University of Wollongong*
Mr Jason Hay *Queensland University of Technology*
Mr Martin J Firth *University of Western Australia*
Dr Scott Salzman *Deakin University*
Ms Charanjit Kaur *Monash University*
Dr Jill Wright *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: Rebecca Pedley, Portfolio Manager; Anna Carter, Development Editor; Julie Ganner, Production Manager and Copy Editor; and Lisa Woodland, Rights & Permissions Team Leader.

how to use this book


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.

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

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.


CHAPTER 1 DEFINING AND COLLECTING DATA 5

LEARNING OBJECTIVES

After studying this chapter you should be able to:

1 identify the types of data used in business
2 identify how statistics is used in business
3 recognise the sources of data used in business
4 distinguish between different survey sampling methods
5 evaluate the quality of surveys

Not so long ago, business students were unfamiliar with the word *data* and had little experience handling data. Today, every time you visit a search engine website or 'ask' your mobile device a question, you are handling data. And if you 'check in' to a location or indicate that you 'like' something, you are *creating* data as well.

You accept as almost true the premises of stories in which characters collect 'a lot of data' to uncover conspiracies, foretell disasters or catch a criminal.

You hear concerns about how the government or business might be able to 'spy' on you in some way or how large social media companies 'mine' your personal data for profit.

You hear the word *data* everywhere and may even have a 'data plan' for your smartphone. You know, in a general way, that data are facts about the world and that most data seem to be, ultimately, a set of numbers – that 34% of students recently polled prefer using a certain Internet browser, or that 50% of citizens believe the country is headed in the right direction, or that unemployment is down 3%, or that your best friend's social media account has 835 friends and 202 recent posts.

You cannot escape from data in this digital world. What, then, should you do? You could try to ignore data and conduct business by relying on hunches or your 'gut instincts'. However, if you want to use only gut instincts, then you probably shouldn't be reading this book or taking business courses in the first place.

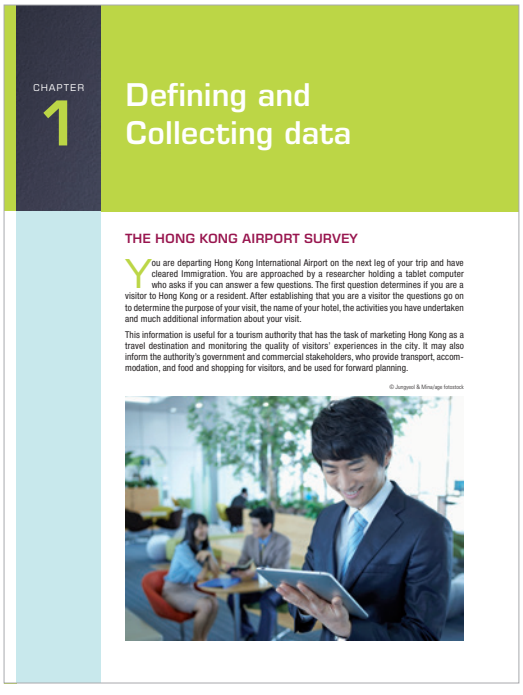
You could note that there is so much data in the world – or just in your own little part of the world – that you couldn't possibly get a handle on it.

You could accept other people's data summaries and their conclusions without first reviewing the data yourself. That, of course, would expose yourself to fraudulent practices.

Or you could do things the proper way and realise the benefits of learning the methods of statistics, the subject of this book. 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 data, or forming conclusions about those data, 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.

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.


CHAPTER 1

Defining and Collecting data

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 & Minoj/age fotostock

Data sets and **Excel workbooks** that accompany the text can be downloaded and used to answer the appropriate questions.

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

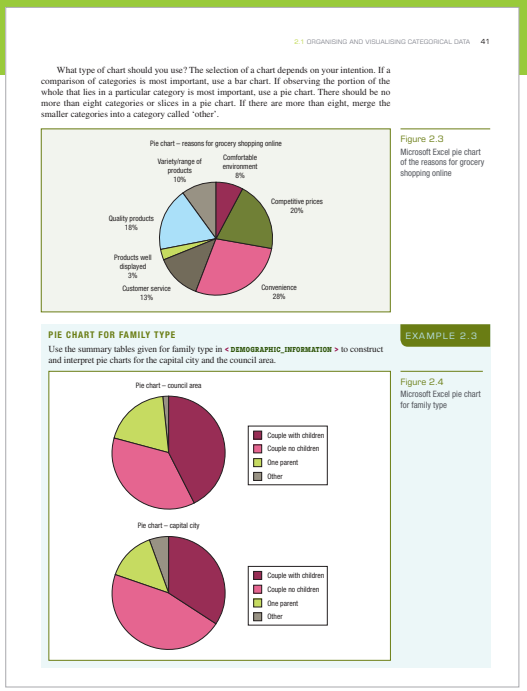
The authors believe that the use of computer software is an integral part of learning statistics. Our focus emphasises analysing data by interpreting the output from Microsoft Excel while reducing emphasis on doing calculations. Excel 2016 changes to statistical functions are reflected in the operations 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*.



2.1 ORGANISING AND VISUALISING CATEGORICAL DATA 41

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. There should be no more than eight categories or slices in a pie chart. If there are more than eight, merge the smaller categories into a category called 'other'.

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

PIE CHART FOR FAMILY TYPE EXAMPLE 2.3

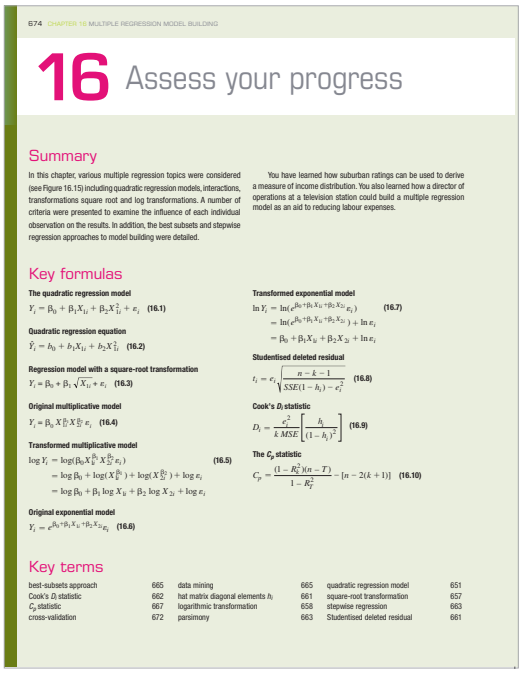
Use the summary tables given for family type in < DEMOGRAPHIC_INFORMATION > to construct and interpret pie charts for the capital city and the council area.

Figure 2.4
Microsoft Excel pie chart for family type

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.



674 CHAPTER 16 MULTIPLE REGRESSION MODEL BUILDING

16 Assess your progress

Summary

In this chapter, various multiple regression topics were considered (see Figure 16.15) including quadratic regression models, interactions, transformations square root and log transformations. A number of criteria were presented to examine the influence of each individual observation on the results. In addition, the best subsets and stepwise regression approaches to model building were detailed.

You have learned how suburban ratings can be used to derive a measure of income distribution. You also learned how a director of operations at a television station could build a multiple regression model as an aid to reducing labour expenses.

Key formulas

The quadratic regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \varepsilon_i \quad (16.1)$$

Quadratic regression equation

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{1i}^2 \quad (16.2)$$

Regression model with a square-root transformation

$$Y_i = \beta_0 + \beta_1 \sqrt{X_{1i}} + \varepsilon_i \quad (16.3)$$

Original multiplicative model

$$Y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} \varepsilon_i \quad (16.4)$$

Transformed multiplicative model

$$\begin{aligned} \log Y_i &= \log(\beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} \varepsilon_i) \\ &= \log \beta_0 + \log(X_{1i}^{\beta_1}) + \log(X_{2i}^{\beta_2}) + \log \varepsilon_i \\ &= \log \beta_0 + \beta_1 \log X_{1i} + \beta_2 \log X_{2i} + \log \varepsilon_i \end{aligned} \quad (16.5)$$

Original exponential model

$$Y_i = e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}} \varepsilon_i \quad (16.6)$$

Transformed exponential model

$$\begin{aligned} \ln Y_i &= \ln(e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}} \varepsilon_i) \\ &= \ln(e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}}) + \ln \varepsilon_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ln \varepsilon_i \end{aligned} \quad (16.7)$$

Studentised deleted residual

$$t_i = e_i \sqrt{\frac{n-k-1}{SSE(1-h_i) - e_i^2}} \quad (16.8)$$

Cook's D_i statistic

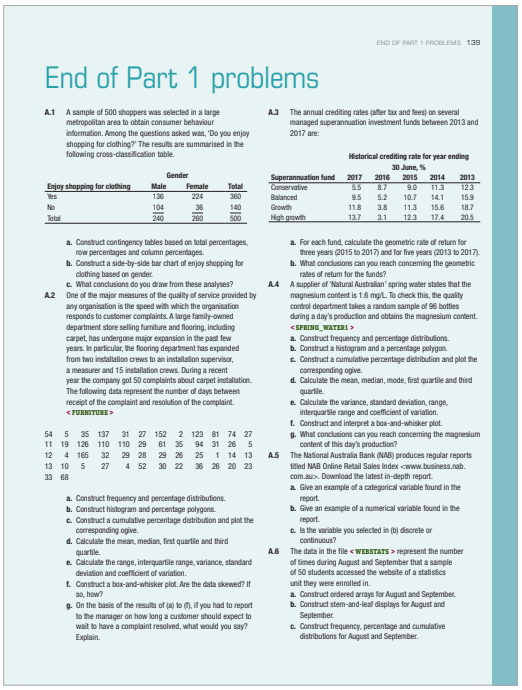
$$D_i = \frac{e_i^2}{k\,MSE}\left[\frac{h_i}{(1-h_i)^2}\right] \quad (16.9)$$

The C_p statistic

$$C_p = \frac{(1-R_k^2)(n-T)}{1-R_T^2} - [n-2(k+1)] \quad (16.10)$$

Key terms

best-subsets approach	665	data mining	665	quadratic regression model	651
Cook's D_i statistic	662	hat matrix diagonal elements h_i	661	square-root transformation	657
C_p statistic	667	logarithmic transformation	658	stepwise regression	663
cross-validation	672	parsimony	663	Studentised deleted residual	661



END OF PART 1 PROBLEMS 139

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.

	Gender		
Enjoy shopping for clothing	Male	Female	Total
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?

A.2 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-owned 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 measurer 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. < FURNITURE >

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	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 range, interquartile range, variance, standard deviation and coefficient of variation.
f. Construct a box-and-whisker plot. Are the data skewed? If so, how?
g. On the basis of the results of (a) to (f), if you had to report to the manager on how long a customer should expect to wait to have a complaint resolved, what would you say? Explain.

A.3 The annual crediting rates (after tax and fees) on several managed superannuation investment funds between 2013 and 2017 are:

	Historical crediting rate for year ending 30 June, %				
Superannuation fund	2017	2016	2015	2014	2013
Conservative	5.5	8.7	9.0	11.3	12.3
Balanced	9.5	5.2	10.7	14.1	15.9
Growth	11.8	3.8	11.3	15.6	18.7
High growth	13.7	3.1	12.3	17.4	20.5

a. For each fund, calculate the geometric rate of return for three years (2015 to 2017) and for five years (2013 to 2017).
b. What conclusions can you reach concerning the geometric rates of return for the funds?

A.4 A supplier of 'Natural Australian' spring water states that the magnesium content is 1.6 mg/L. To check this, the quality control department takes a random sample of 96 bottles during a day's production and obtains the magnesium content. < SPRING_WATER1 >

a. Construct frequency and percentage distributions.
b. Construct a histogram and a percentage polygon.
c. Construct a cumulative percentage distribution and plot the corresponding ogive.
d. Calculate the mean, median, mode, 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 National Australia Bank (NAB) produces regular reports titled NAB Online Retail Sales Index <www.business.nab.com.au>. Download the latest in-depth report.

a. Give an example of a categorical variable found in the report.
b. Give an example of a numerical variable found in the report.
c. Is the variable you selected in (b) discrete or continuous?

A.6 The data in the file < WEBSTATS > represent the number of times during August and September that a sample of 50 students accessed the website of a statistics unit they were enrolled in.

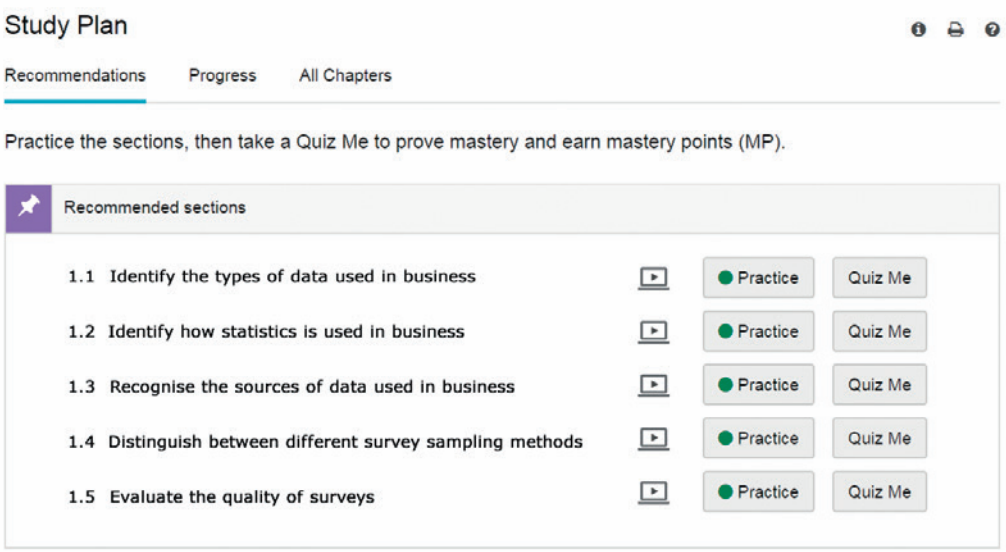
a. Construct ordered arrays for August and September.
b. Construct stem-and-leaf displays for August and September.
c. Construct frequency, percentage and cumulative distributions for August and September.

*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.

MyLab Statistics
a guided tour for students and educators

Study Plan

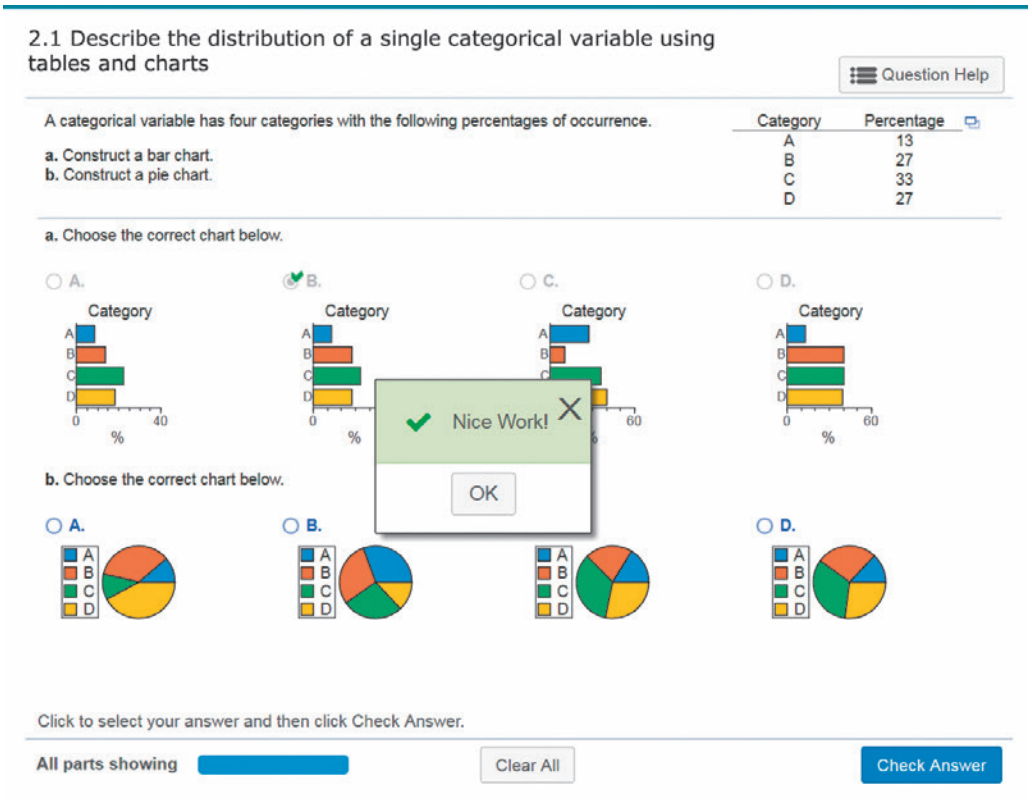
A study plan is generated from each student's results on a pre-test. Students can clearly see which topics they have mastered and, more importantly, which they need to work on.

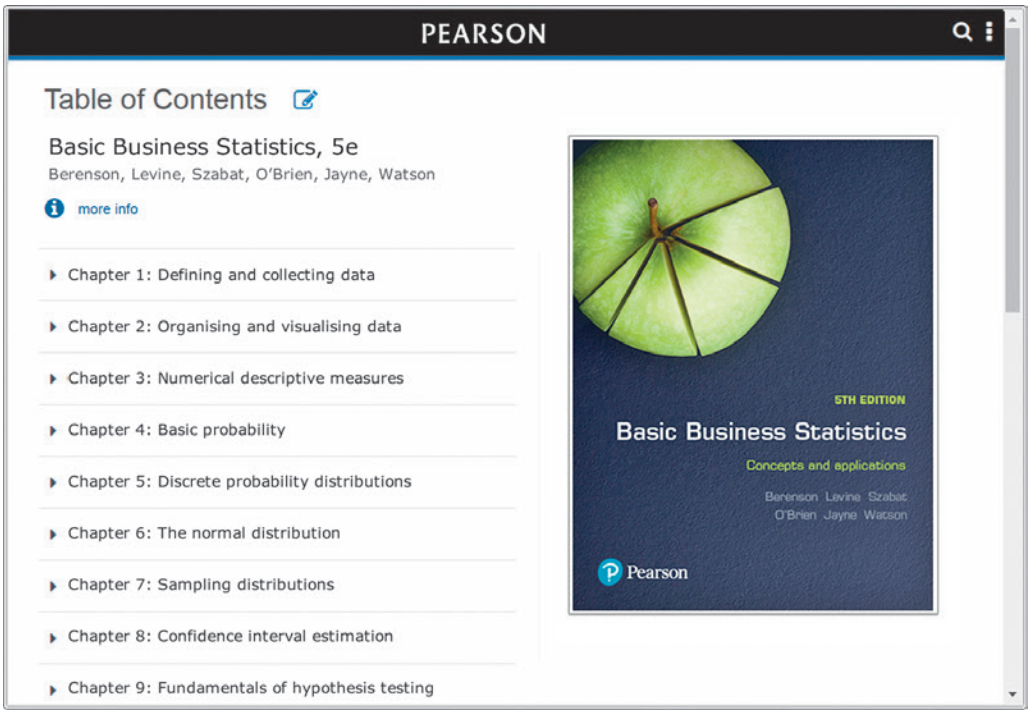


Unlimited Practice

Each MyLab Statistics 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 though study plan or homework exercises, instant feedback and tutorial resources guide them towards understanding.

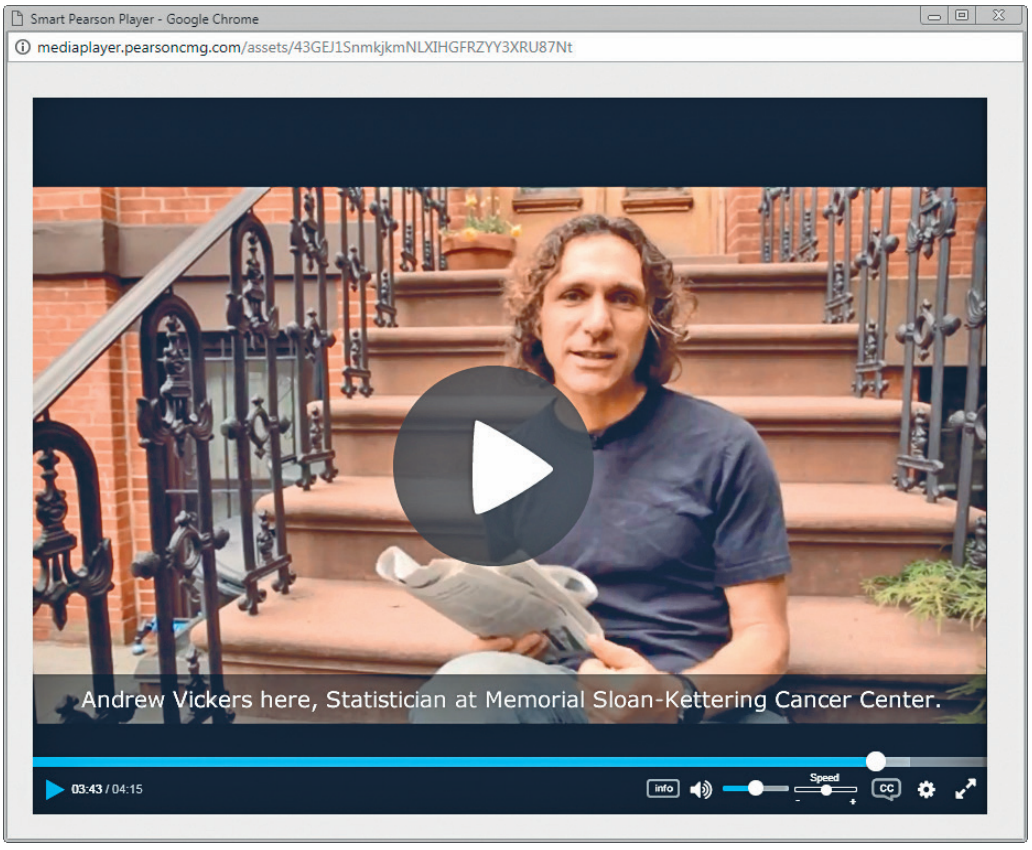




Learning Resources

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.



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 is 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. Questions can be integrated into Blackboard, Canvas or Moodle Learning Management Systems.

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 postgraduate 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 and Teaching in 2013. 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.

Nicola's 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 in economics, Director of the Centre for Human and Social Capital Research, and Director of the MBA program in the Sydney Business School, University of Wollongong. Martin earned his Bachelor of Commerce (first-class honours) and PhD in Economics at the University of Newcastle. His 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. Martin has been an expert witness for a number of Fair Work Commission cases, providing statistical analyses of the effects of penalty rates, workforce casualisation and family and domestic violence leave.

Martin has taught a wide range of quantitative subjects at university level, including business statistics, business analytics, quantitative analysis for decision making, econometrics, financial modelling and business research methods. He also has a keen interest in learning analytics and the development and analysis of new teaching technologies.

about 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, an MBA in business statistics from City College of New York and a PhD in business from the City University of New York. His 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 nationally recognised 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.*

He also is 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 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.

a quick q&a

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.

CHAPTER

1

Defining and Collecting data

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.

LEARNING OBJECTIVES

After studying this chapter you should be able to:

1. identify the types of data used in business
2. identify how statistics is used in business
3. recognise the sources of data used in business
4. distinguish between different survey sampling methods
5. evaluate the quality of surveys

Not so long ago, business students were unfamiliar with the word *data* and had little experience handling data. Today, every time you visit a search engine website or 'ask' your mobile device a question, you are handling data. And if you 'check in' to a location or indicate that you 'like' something, you are *creating* data as well.

You accept as almost true the premises of stories in which characters collect 'a lot of data' to uncover conspiracies, foretell disasters or catch a criminal.

You hear concerns about how the government or business might be able to 'spy' on you in some way or how large social media companies 'mine' your personal data for profit.

You hear the word *data* everywhere and may even have a 'data plan' for your smartphone. You know, in a general way, that data are facts about the world and that most data seem to be, ultimately, a set of numbers – that 34% of students recently polled prefer using a certain Internet browser, or that 50% of citizens believe the country is headed in the right direction, or that unemployment is down 3%, or that your best friend's social media account has 835 friends and 202 recent posts.

You cannot escape from data in this digital world. What, then, should you do? You could try to ignore data and conduct business by relying on hunches or your 'gut instincts'. However, if you want to use only gut instincts, then you probably shouldn't be reading this book or taking business courses in the first place.

You could note that there is so much data in the world – or just in your own little part of the world – that you couldn't possibly get a handle on it.

You could accept other people's data summaries and their conclusions without first reviewing the data yourself. That, of course, would expose yourself to fraudulent practices.

Or you could do things the proper way and realise the benefits of learning the methods of statistics, the subject of this book. 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 data, or forming conclusions about those data, 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 DATA AND STATISTICS

The Meaning of 'Data'

What do we mean by the word *data*? Its common use is somewhat different from its use in statistics. It could be described in a general way as meaning 'facts about the world'. However, statisticians distinguish between the traits or properties that relate to people or things and the actual values that these take.

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

data
The observed values of variables.

VARIABLES

Variables are characteristics of items or individuals.

DATA

Data are the observed values of variables.

For a group of people, we could examine the traits of age, country of birth or weight. For a group of cars, we could note the colour, current value or kilometres driven. These characteristics are called **variables**.

Data are the values associated with these traits or properties. As an example, in Table 1.1 we find a set of data collected from six people which represents observations on three different variables.

Table 1.1

Variable	Data
Age in years	24, 18, 53, 16, 22, 31
Country of birth	Australia, China, Australia, Malaysia, India, Australia
Weight in kilograms	50.2, 74.6, 96.3, 45.2, 56.1, 87.3

In this book, the word *data* is always plural to remind you that data are a collection or set of values. While we could say that a single value, such as 'Australia' is a *datum*, the terms *data point, observation, response* or *single data value* are more typically encountered.

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 (ABS). The ABS needs to collect information about the country of birth of a person and also the countries in which their father and mother were born. While this might seem straightforward, definitional problems arise in the case of people who were adopted or have step- or foster parents or other guardians. So the operational definition used is:

- 'Country of birth of person', which is the country identified as being the one in which the person was born
- 'Country of birth of father', which is the country in which the person's birth father was born, and
- 'Country of birth of mother', which is the country in which the person's birth mother was born (Australian Bureau of Statistics, *Country of Birth Standard,* Cat. No. 1200.0.55.004, 2016).

The Meaning of 'Statistics'

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

Statistics is the branch of mathematics that examines ways to process and analyse data. It provides procedures to collect and transform data in ways that are useful to business decision makers.

Statistics allows you to determine whether your data represent information that could be used in making better decisions. Therefore, it helps you determine whether differences in the

numbers are meaningful in a significant way or are due to chance. To illustrate, consider the following reports:

- In 'News use across social media platforms 2016' the Pew Research Center reported in May 2016, that 67% of the adult US population had a Facebook account and 66% of users get news from the site (<http://assets.pewresearch.org/wpcontent/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf>, accessed 12 June 2017).
- In a blog titled 'The top 10 benefits of newspaper advertising', the 360 Degree Marketing Group says that a study showed newspaper advertising was considered a more trusted paid medium for information (58%) compared with television (54%), radio (49%) or online (27%) (<www.360degreemarketing.com.au/Blog/bid/407663/The-Top-10-Benefits-of-Newspaper-Advertising>, accessed 12 June 2017).

Without statistics, you cannot determine whether the 'numbers' in these stories represent useful information. Without statistics, you cannot validate claims such as the statement that advertising in newspapers or on television is more trusted than online advertising. And without statistics, you cannot see patterns that large amounts of data sometimes reveal.

Statistics is a way of thinking that can help you make better decisions. It helps you solve problems that involve decisions based on data that have been collected. You may have had some statistics instruction in the past. If you ever created a chart to summarise data or calculated values such as averages to summarise data, you have used statistics. But there's even more to statistics than these commonly taught techniques, as the detailed table of contents shows.

Statistics is undergoing important changes today. There are new ways of visualising data that did not exist, were not practicable or were not widely known until recently. And, increasingly, statistics today is being used to 'listen' to what the data might be telling you rather than just being a way to use data to prove something you want to say.

If you associate statistics with doing a lot of mathematical calculations, you will quickly learn that business statistics uses software to perform the calculations for you (and, generally, the software calculates with more precision and efficiency than you could do manually). But while you do not need to be a good manual calculator to apply statistics, because statistics is a way of thinking, you do need to follow a framework or plan to minimise possible errors of thinking and analysis.

One such framework consists of the following tasks to help apply statistics to business decision making:

1. **Define** the data that you want to study in order to solve a problem or meet an objective.
2. **Collect** the data from appropriate sources.
3. **Organise** the data collected by developing tables.
4. **Visualise** the data collected by developing charts.
5. **Analyse** the data collected to reach conclusions and present those results.

Typically, you do the tasks in the order listed. You must always do the first two tasks to have meaningful outcomes, but, in practice, the order of the other three can change or appear inseparable. Certain ways of visualising data will help you to organise your data while performing preliminary analysis as well. In any case, when you apply statistics to decision making, you should be able to identify all five tasks, and you should verify that you have done the first two tasks before the other three.

Using this framework helps you to apply statistics to these four broad categories of business activities:

1. Summarise and visualise business data.
2. Reach conclusions from those data.
3. Make reliable forecasts about business activities.
4. Improve business processes.

Throughout this book, and especially in the scenarios that begin the chapters, you will discover specific examples of how we can apply statistics to business situations.

Statistics is itself 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
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.

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 Australian Census held in 2016 was badly affected by a computer shutdown on Census night, 9 August. It was blamed on the need to protect the system from denial of service cyber attacks and added approximately $30 million to the cost of the Census (<www.abc.net.au/ news/2016-10-25/turning-router-off-and-on-could-have-prevented-census-outage/7963916>, accessed 13 July 2017).

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.

Other Important Definitions

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

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.

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.

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.

1.2 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.

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?