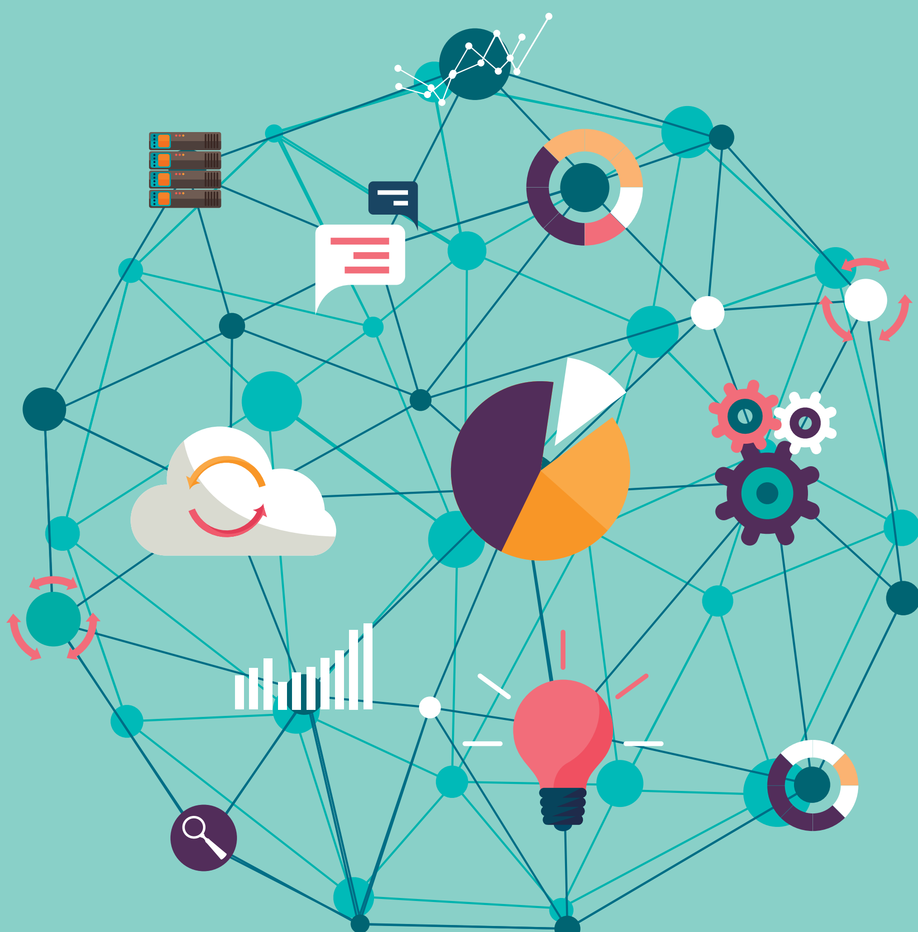
Seventh Edition

STATISTICS FOR ECONOMICS, Accounting and Business Studies

MICHAEL BARROW



Pearson

Statistics for Economics,
Accounting and Business Studies

Statistics for Economics, Accounting and Business Studies

Seventh edition

Michael Barrow

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Pearson Education Limited
Edinburgh Gate
Harlow CM20 2JE
United Kingdom
Tel: +44 (0)1279 623623
Web: www.pearson.com/uk

First published 1988 (print)
Second edition published 1996 (print)
Third edition published 2001 (print and electronic)
Fourth edition published 2006 (print and electronic)
Fifth edition published 2009 (print and electronic)
Sixth edition published 2013 (print and electronic)
Seventh edition published 2017 (print and electronic)

ISBN: 978-1-292-11870-3 (Print)
978-1-292-11874-1 (PDF)
978-1-292-18249-0 (ePub)

British Library Cataloguing-in-Publication Data
A catalogue record for the print edition is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Barrow, Michael, author.
Title: Statistics for economics, accounting and business studies / Michael Barrow.
Description: Seventh edition. | Harlow, United Kingdom : Pearson Education, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2016049343 | ISBN 9781292118703 (Print) | ISBN 9781292118741 (PDF) | ISBN 9781292182490 (ePub)
Subjects: LCSH: Economics--Statistical methods. | Commercial statistics.
Classification: LCC HB137 .B37 2016 | DDC 519.5024/33--dc23
LC record available at https://lccn.loc.gov/2016049343

10 9 8 7 6 5 4 3 2 1
21 20 19 18 17

Print edition typeset in 9/12pt StoneSerifITCPro-Medium by iEnergizer Aptara® Ltd
Printed in Slovakia by Neografia

NOTE THAT ANY PAGE CROSS REFERENCES REFER TO THE PRINT EDITION

For Patricia, Caroline and Nicolas

This page intentionally left blank

Contents

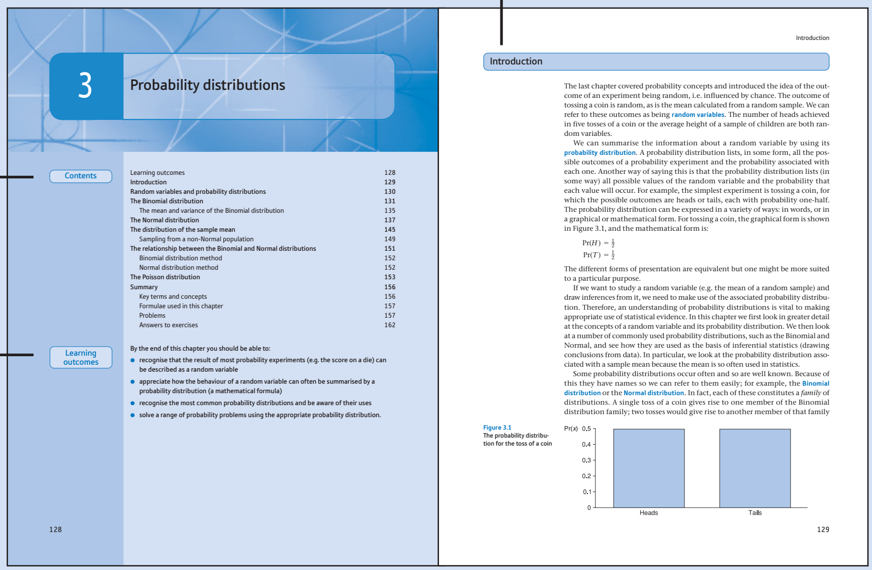
Guided tour of the book

Setting the scene

Chapter introductions set the scene for learning and link the chapters together.

Chapter contents guide you through the chapter, highlighting key topics and showing you where to find them.

Learning outcomes summarise what you should have learned by the end of the chapter.



3 Probability distributions

Contents

Learning outcomes	128
Introduction	**129**
Random variables and probability distributions	**130**
The Binomial distribution	**131**
The mean and variance of the Binomial distribution	135
The Normal distribution	**137**
The distribution of the sample mean	**145**
Sampling from a non-Normal population	149
The relationship between the Binomial and Normal distributions	**151**
Binomial distribution method	152
Normal distribution method	152
The Poisson distribution	**153**
Summary	**156**
Key terms and concepts	156
Formulae used in this chapter	157
Problems	157
Answers to exercises	162

Learning outcomes

By the end of this chapter you should be able to:

- recognise that the result of most probability experiments (e.g. the score on a die) can be described as a random variable
- appreciate how the behaviour of a random variable can often be summarised by a probability distribution (a mathematical formula)
- recognise the most common probability distributions and be aware of their uses
- solve a range of probability problems using the appropriate probability distribution.

128

Introduction

Introduction

The last chapter covered probability concepts and introduced the idea of the outcome of an experiment being random, i.e. influenced by chance. The outcome of tossing a coin is random, as is the mean calculated from a random sample. We can refer to these outcomes as being **random variables**. The number of heads achieved in five tosses of a coin or the average height of a sample of children are both random variables.

We can summarise the information about a random variable by using its **probability distribution**. A probability distribution lists, in some form, all the possible outcomes of a probability experiment and the probability associated with each one. Another way of saying this is that the probability distribution lists (in some way) all possible values of the random variable and the probability that each value will occur. For example, the simplest experiment is tossing a coin, for which the possible outcomes are heads or tails, each with probability one-half. The probability distribution can be expressed in a variety of ways: in words, or in a graphical or mathematical form. For tossing a coin, the graphical form is shown in Figure 3.1, and the mathematical form is:

$$\Pr(H) = \tfrac{1}{2}$$
$$\Pr(T) = \tfrac{1}{2}$$

The different forms of presentation are equivalent but one might be more suited to a particular purpose.

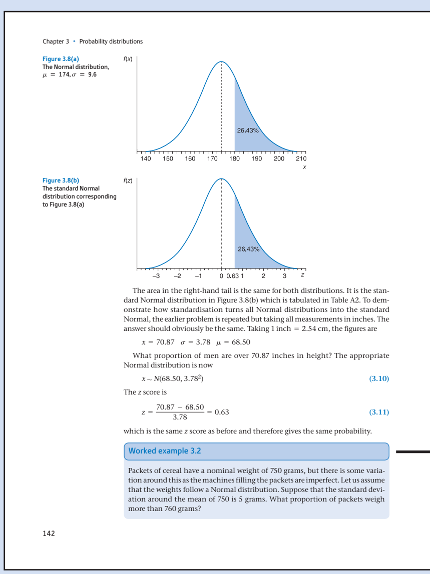
If we want to study a random variable (e.g. the mean of a random sample) and draw inferences from it, we need to make use of the associated probability distribution. Therefore, an understanding of probability distributions is vital to making appropriate use of statistical evidence. In this chapter we first look in greater detail at the concepts of a random variable and its probability distribution. We then look at a number of commonly used probability distributions, such as the Binomial and Normal, and see how they are used as the basis of inferential statistics (drawing conclusions from data). In particular, we look at the probability distribution associated with a sample mean because the mean is so often used in statistics.

Some probability distributions occur often and so are well known. Because of this they have names so we can refer to them easily; for example, the **Binomial distribution** or the **Normal distribution**. In fact, each of these constitutes a *family* of distributions. A single toss of a coin gives rise to one member of the Binomial distribution family; two tosses would give rise to another member of that family

Figure 3.1
The probability distribution for the toss of a coin

129

Practising and testing your understanding



Chapter 3 • Probability distributions

Figure 3.8(a)
The Normal distribution, $\mu = 174, \sigma = 9.6$

Figure 3.8(b)
The standard Normal distribution corresponding to Figure 3.8(a)

The area in the right-hand tail is the same for both distributions. It is the standard Normal distribution in Figure 3.8(b) which is tabulated in Table A2. To demonstrate how standardisation turns all Normal distributions into the standard Normal, the earlier problem is repeated but taking all measurements in inches. The answer should obviously be the same. Taking 1 inch = 2.54 cm, the figures are

$$x = 70.87 \quad \sigma = 3.78 \quad \mu = 68.50$$

What proportion of men are over 70.87 inches in height? The appropriate Normal distribution is now

$$x \sim N(68.50, 3.78^2) \qquad (3.10)$$

The z score is

$$z = \frac{70.87 - 68.50}{3.78} = 0.63 \qquad (3.11)$$

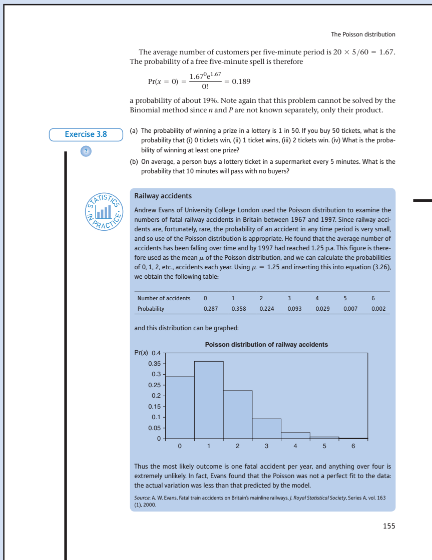
which is the same z score as before and therefore gives the same probability.

Worked example 3.2

Packets of cereal have a nominal weight of 750 grams, but there is some variation around this as the machines filling the packets are imperfect. Let us assume that the weights follow a Normal distribution. Suppose that the standard deviation around the mean of 750 is 5 grams. What proportion of packets weigh more than 760 grams?

142

Worked examples break down statistical techniques step-by-step and illustrate how to apply an understanding of statistical techniques to real life.



The Poisson distribution

The average number of customers per five-minute period is $20 \times 5/60 = 1.67$. The probability of a free five-minute spell is therefore

$$\Pr(x = 0) = \frac{1.67^0 e^{1.67}}{0!} = 0.189$$

a probability of about 19%. Note again that this problem cannot be solved by the Binomial method since n and P are not known separately, only their product.

Exercise 3.8

(a) The probability of winning a prize in a lottery is 1 in 50. If you buy 50 tickets, what is the probability that (i) 0 tickets win, (ii) 1 ticket wins, (iii) 2 tickets win. (iv) What is the probability of winning at least one prize?

(b) On average, a person buys a lottery ticket in a supermarket every 5 minutes. What is the probability that 10 minutes will pass with no buyers?

STATISTICS IN PRACTICE

Railway accidents

Andrew Evans of University College London used the Poisson distribution to examine the numbers of fatal railway accidents in Britain between 1967 and 1997. Since railway accidents are, fortunately, rare, the probability of an accident in any time period is very small, and so use of the Poisson distribution is appropriate. He found that the average number of accidents has been falling over time and by 1997 had reached 1.25 p.a. This figure is therefore used as the mean μ of the Poisson distribution, and we can calculate the probabilities of 0, 1, 2, etc., accidents each year. Using $\mu = 1.25$ and inserting this into equation (3.26), we obtain the following table:

Number of accidents	0	1	2	3	4	5	6
Probability	0.287	0.358	0.224	0.093	0.029	0.007	0.002

and this distribution can be graphed:

Poisson distribution of railway accidents

Pr(x) 0.4 0.35 0.3 0.25 0.2 0.15 0.1 0.05 0

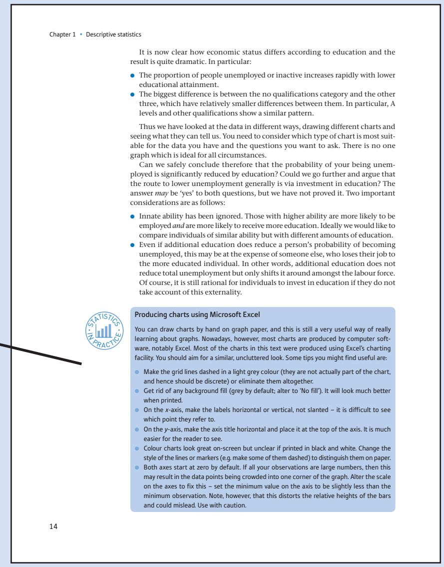
0 1 2 3 4 5 6

Thus the most likely outcome is one fatal accident per year, and anything over four is extremely unlikely. In fact, Evans found that the Poisson was not a perfect fit to the data: the actual variation was less than that predicted by the model.

Source: A. W. Evans, Fatal train accidents on Britain's mainline railways, *J. Royal Statistical Society*, Series A, vol. 163 (1), 2000.

155

Statistics in practice provide real and interesting applications of statistical techniques in business practice.
They also provide helpful hints on how to use different software packages such as Excel and calculators to solve statistical problems and help you manipulate data.



Chapter 1 • Descriptive statistics

It is now clear how economic status differs according to education and the result is quite dramatic. In particular:

- The proportion of people unemployed or inactive increases rapidly with lower educational attainment.
- The biggest difference is between the no qualifications category and the other three, which have relatively smaller differences between them. In particular, A levels and other qualifications show a similar pattern.

Thus we have looked at the data in different ways, drawing different charts and seeing what they can tell us. You need to consider which type of chart is most suitable for the data you have and the questions you want to ask. There is no one graph which is ideal for all circumstances.

Can we safely conclude therefore that the probability of your being unemployed is significantly reduced by education? Could we go further and argue that the route to lower unemployment generally is via investment in education? The answer *may* be 'yes' to both questions, but we have not proved it. Two important considerations are as follows:

- Innate ability has been ignored. Those with higher ability are more likely to be employed *and* are more likely to receive more education. Ideally we would like to compare individuals of similar ability but with different amounts of education.
- Even if additional education does reduce a person's probability of becoming unemployed, this may be at the expense of someone else, who loses their job to the more educated individual. In other words, additional education does not reduce total unemployment but only shifts it around amongst the labour force. Of course, it is still rational for individuals to invest in education if they do not take account of this externality.

STATISTICS IN PRACTICE

Producing charts using Microsoft Excel

You can draw charts by hand on graph paper, and this is still a very useful way of really learning about graphs. Nowadays, however, most charts are produced by computer software, notably Excel. Most of the charts in this text were produced using Excel's charting facility. You should aim for a similar, uncluttered look. Some tips you might find useful are:

- Make the grid lines dashed in a light grey colour (they are not actually part of the chart, and hence should be discrete) or eliminate them altogether.
- Get rid of any background fill (grey by default; alter to 'No fill'). It will look much better when printed.
- On the *x*-axis, make the labels horizontal or vertical, not slanted – it is difficult to see which point they refer to.
- On the *y*-axis, make the axis title horizontal and place it at the top of the axis. It is much easier for the reader to see.
- Colour charts look great on-screen but unclear if printed in black and white. Change the style of the lines or markers (e.g. make some of them dashed) to distinguish them on paper.
- Both axes start at zero by default. If all your observations are large numbers, then this may result in the data points being crowded into one corner of the graph. Alter the scale on the axes to fix this – set the minimum value on the axis to be slightly less than the minimum observation. Note, however, that this distorts the relative heights of the bars and could mislead. Use with caution.

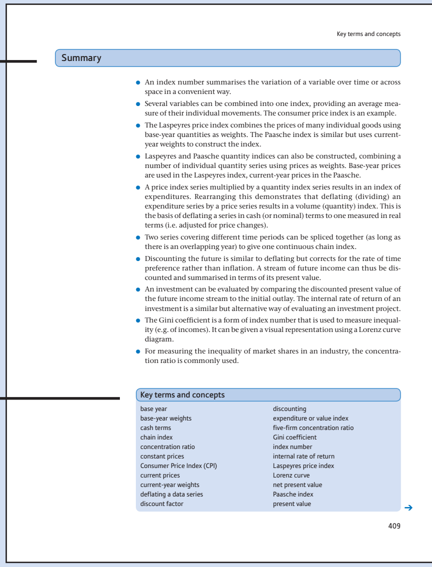
14

Exercises throughout the chapter allow you to stop and check your understanding of the topic you have just learned. You can check the **answers** at the end of each chapter.

Reinforcing your understanding

Problems at the end of each chapter range in difficulty to provide a more in-depth practice of topics.

Chapter summaries recap all the important topics covered in the chapter.



Key terms and concepts

Summary

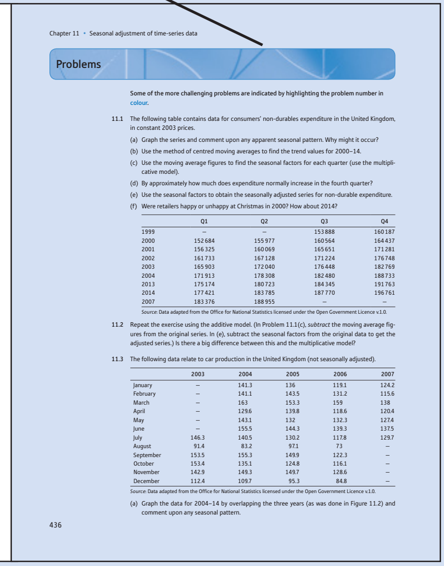
- An index number summarises the variation of a variable over time or across space in a convenient way.
- Several variables can be combined into one index, providing an average measure of their individual movements. The consumer price index is an example.
- The Laspeyres price index combines the prices of many individual goods using base-year quantities as weights. The Paasche index is similar but uses current-year weights to construct the index.
- Laspeyres and Paasche quantity indices can also be constructed, combining a number of individual quantity series using prices as weights. Base-year prices are used in the Laspeyres index, current-year prices in the Paasche.
- A price index series multiplied by a quantity index series results in an index of expenditures. Rearranging this demonstrates that deflating (dividing) an expenditure series by a price series results in a volume (quantity) index. This is the basis of deflating a series in cash (or nominal) terms to one measured in real terms (i.e. adjusted for price changes).
- Two series covering different time periods can be spliced together (as long as there is an overlapping year) to give one continuous chain index.
- Discounting the future is similar to deflating but corrects for the rate of time preference rather than inflation. A stream of future income can thus be discounted and summarised in terms of its present value.
- An investment can be evaluated by comparing the discounted present value of the future income stream to the initial outlay. The internal rate of return of an investment is a similar but alternative way of evaluating an investment project.
- The Gini coefficient is a form of index number that is used to measure inequality (e.g. of incomes). It can be given a visual representation using a Lorenz curve diagram.
- For measuring the inequality of market shares in an industry, the concentration ratio is commonly used.

Key terms and concepts

base year	discounting
base-year weights	expenditure or value index
cash terms	five-firm concentration ratio
chain index	Gini coefficient
concentration ratio	index number
constant prices	internal rate of return
Consumer Price Index (CPI)	Laspeyres price index
current prices	Lorenz curve
current-year weights	net present value
deflating a data series	Paasche index
discount factor	present value

→

409



Chapter 11 • Seasonal adjustment of time-series data

Problems

Some of the more challenging problems are indicated by highlighting the problem number in colour.

11.1 The following table contains data for consumers' non-durables expenditure in the United Kingdom, in constant 2003 prices.

(a) Graph the series and comment upon any apparent seasonal pattern. Why might it occur?

(b) Use the method of centred moving averages to find the trend values for 2000–14.

(c) Use the moving average figures to find the seasonal factors for each quarter (use the multiplicative model).

(d) By approximately how much does expenditure normally increase in the fourth quarter?

(e) Use the seasonal factors to obtain the seasonally adjusted series for non-durable expenditure.

(f) Were retailers happy or unhappy at Christmas in 2000? How about 2014?

	Q1	Q2	Q3	Q4
1999	–	–	153 888	160 187
2000	152 684	155 977	160 564	164 437
2001	156 325	160 069	165 651	171 281
2002	161 733	167 128	171 224	176 748
2003	165 903	172 040	176 448	182 769
2004	171 913	178 308	182 480	188 733
2013	175 174	180 723	184 345	191 763
2014	177 421	183 785	187 770	196 761
2007	183 376	188 955	–	–

Source: Data adapted from the Office for National Statistics licensed under the Open Government Licence v.1.0.

11.2 Repeat the exercise using the additive model. (In Problem 11.1(c), *subtract* the moving average figures from the original series. In (e), subtract the seasonal factors from the original data to get the adjusted series.) Is there a big difference between this and the multiplicative model?

11.3 The following data relate to car production in the United Kingdom (not seasonally adjusted).

	2003	2004	2005	2006	2007
January	–	141.3	136	119.1	124.2
February	–	141.1	143.5	131.2	115.6
March	–	163	153.3	159	138
April	–	129.6	139.8	118.6	120.4
May	–	143.1	132	132.3	127.4
June	–	155.5	144.3	139.3	137.5
July	146.3	140.5	130.2	117.8	129.7
August	91.4	83.2	97.1	73	–
September	153.5	155.3	149.9	122.3	–
October	153.4	135.1	124.8	116.1	–
November	142.9	149.3	149.7	128.6	–
December	112.4	109.7	95.3	84.8	–

Source: Data adapted from the Office for National Statistics licensed under the Open Government Licence v.1.0.

(a) Graph the data for 2004–14 by overlapping the three years (as was done in Figure 11.2) and comment upon any seasonal pattern.

436

Key terms and concepts are highlighted when they first appear in the text and are brought together at the end of each chapter.

Publisher's acknowledgements

We are grateful to the following for permission to reproduce copyright material:

Figures

Figure on page 22 from US Census Bureau. World Population Profile 2000, US Census Bureau/www.unaids.org; Figure on page 71 from Fig 1.26 CO_2 emissions versus real GDP in 1950, Gapminder Foundation; Figure on page 71 from Figure 1.27 CO_2 emissions versus real GDP in 2008, Gapminder Foundation; Figure on page 297 from Powered by Trendalyzer, 1981, Gapminder Foundation; Figure on page 342 from R. Dornbusch and S. Fischer (in R.E. Caves and L.B. Krause), *Britain's Economic Performance*, Brookings, 1980, Brookings Institution Press; Figure on page 352 adapted from the 1999 edition of *Economic Trends Annual Supplement*, Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0; Figure on page 408 from T. Beck et al., Bank concentration and fragility: impact and dynamics. *NBER Working Paper 11500*, © 2005 by Thorsten Beck, AsliDemirgüç-Kunt and Ross Levine. All rights reserved.

Screenshots

Screenshots on page 41, page 244, page 361, page 397 from Microsoft Corporation, Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.

Tables

Table on page 11 adapted from Department for Children, Schools and Families, *Education and Training Statistics for the UK 2009*, contains public sector information licensed under the Open Government Licence v2.0; Table on page 16 adapted from data from the Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 17 adapted from *HM Revenue and Customs Statistics*, 2005, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 25 from The Economics Network, University of Bristol – Economics Network Team; Table on page 48 adapted from data from the Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 57 adapted from World Bank, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 66 adapted from Greece records lowest life satisfaction rating of all OECD countries, *The Guardian*, 01/07/2015 and Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 66 adapted from Organisation for Economic Co-Operation and Development and Office for National Statistics, Contains public sector information licensed under the Open

Government Licence v2.0; Table on page 238 adapted from the UK government's transport data, Contains public sector information licensed under the Open Government Licence v2.0; Table on pages 363–4 adapted from *The UK Time Use Survey*, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 378 adapted from *The Digest of UK Energy Statistics*, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 384 from *The Human Development Index*, 1980–2013,United Nations Development Programme, Creative Commons Attribution license (CC BY 3.0 IGO); Table on page 399 adapted from *The Family Resources Survey 2006–07*, published by the Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 404 adapted from *The Effects of Taxes and Benefits on Household Income*, 2009/10, Office of National Statistics, 2011, Contains public sector information licensed under the Open Government Licence v2.0; Table 10.25 on page 405 from *Long-Run Changes in British Income Equality*, Soltow L. (2008), © John Wiley and Sons; Table 10.26 on page 405 from Real GDP per capita for more than one hundred countries, *Economic Journal*, vol. 88, I.B. Kravis, A.W Heston, R. Summers, 1978, Organisation for Economic Co-operation and Development (OECD); Table 10.27 on page 407 from *World Development Report*, 2006, © World Bank, Creative Commons Attribution license (CC BY 3.0 IGO); Table 10.28 on page 407 adapted from National Archives, Contains public sector information licensed under the Open Government Licence v2.0; Table on page 422 adapted from data from the Office for National Statistics, *UK unemployed aged over 16 – not seasonally adjusted*, Contains public sector information licensed under the Open Government Licencev2.0; Tables on page 436 adapted from Data from the Office for National Statistics, Contains public sector information licensed under the Open Government Licence v2.0.

Text

Article on page 59 adapted from How a $1 investment can grow over time, The Economist, 12 February 2000, republished with permission of Economist Newspaper Group; Extract on page 155 from A.W. Evans, Fatal train accidents on Britain's mainline railways, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2000, © John Wiley and Sons; Extract on pages 176–7 from Music down the phone, *The Times*, 10/07/2000, © News Syndication; Exercise on page 204 adapted from *Statistical Inference: Commentary for the Social and Behavioral Sciences*, W. Oakes, 1986, reproduced with permission of John Wiley & Sons, Inc.; Activity on page 208 from Do children prefer branded goods only because of the name?, *The New Scientist*, © 2007 Reed Business Information UK , all rights reserved, distributed by Tribune Content Agency; Article on page 245 from J. Ermisch and M. Francesconi, Cohabitation in Great Britain: not for long, but here to stay, *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 2002, © John Wiley and Sons; Extract on page 265 from *Economic Development in the Third World*, Pearson Education (Todaro, M. 1992) and © World Bank. Creative Commons Attribution license (CC BY 3.0 IGO); Extract on page 319 adapted from World Bank, © World Bank. Creative Commons Attribution license (CC BY 3.0 IGO); Article on page 369 from M. Collins, Editorial: Sampling for UK telephone surveys, *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 2002, © John Wiley and Sons.

Preface to the seventh edition

This text is aimed at students of economics and the closely related disciplines of accountancy, finance and business, and provides examples and problems relevant to those subjects, using real data where possible. The text is at a fairly elementary university level and requires no prior knowledge of statistics, nor advanced mathematics. For those with a weak mathematical background and in need of some revision, some recommended texts are given at the end of this preface.

This is not a cookbook of statistical recipes: it covers all the relevant concepts so that an understanding of why a particular statistical technique should be used is gained. These concepts are introduced naturally in the course of the text as they are required, rather than having sections to themselves. The text can form the basis of a one- or two-term course, depending upon the intensity of the teaching.

As well as explaining statistical concepts and methods, the different schools of thought about statistical methodology are discussed, giving the reader some insight into some of the debates that have taken place in the subject. The text uses the methods of classical statistical analysis, for which some justification is given in Chapter 5, as well as presenting criticisms that have been made of these methods.

Changes in this edition

There are limited changes in this edition, apart from a general updating of the examples used in the text. Other changes include:

- A new section on how to write statistical reports (Chapter 1)
- Examples of good and bad graphs, and how to improve them
- Illustrations of graphing regression coefficients as a means of presentation
- Probability chapter expanded to improve exposition
- More discussion and critique of hypothesis testing
- New Companion Website for students including quizzes to test your knowledge and Excel data files
- As before, there is an associated blog on statistics and the teaching of the subject. This is where I can comment on interesting stories and statistical issues, relating them to topics covered in this text. You are welcome to comment on the posts and provide feedback on the text. The blog can be found at http://anecdotesandstatistics.blogspot.co.uk/.

  For lecturers:
  - As before, PowerPoint slides are available containing most of the key tables, formulae and diagrams, which can be adapted for lecture use
  - Answers to even-numbered problems (not included in the text itself)
  - An Instructor's Manual giving hints and guidance on some of the teaching issues, including those that come up in response to some of the exercises and problems.

For students:

- The associated website contains numerous exercises (with answers) for the topics covered in this text. Many of these contained randomised values so that you can try out the tests several times and keep track of you progress and understanding.

Mathematics requirements and suggested texts

No more than elementary algebra is assumed in this text, any extensions being covered as they are needed in the text. It is helpful to be comfortable with manipulating equations, so if some revision is needed, I recommend one of the following books:

Jacques, I., *Mathematics for Economics and Business*, 8th edn, Pearson, 2015
Renshaw, G., *Maths for Economists*, 4th edn, Oxford University Press, 2016.

Acknowledgements

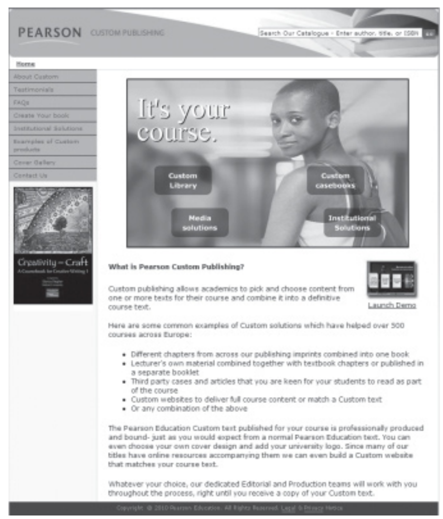
I would like to thank the reviewers who made suggestions for this new edition and to the many colleagues and students who have passed on comments or pointed out errors or omissions in previous editions. I would like to thank the editors at Pearson, especially Caitlin Lisle and Carole Drummond, who have encouraged me, responded to my various queries and gently reminded me of impending deadlines. I would also like to thank my family for giving me encouragement and time to complete this edition.

Custom publishing

Custom publishing allows academics to pick and choose content from one or more textbooks for their course and combine it into a definitive course text.

Here are some common examples of custom solutions which have helped over 800 courses across Europe:

- different chapters from across our publishing imprints combined into one book;
- lecturer's own material combined together with textbook chapters or published in a separate booklet;
- third-party cases and articles that you are keen for your students to read as part of the course;
- any combination of the above.



The Pearson custom text published for your course is professionally produced and bound – just as you would expect from a normal Pearson text. Since many of our titles have online resources accompanying them we can even build a Custom website that matches your course text.

If you are teaching an introductory statistics course for economics and business students, do you also teach an introductory mathematics course for economics and business students? If you do, you might find chapters from *Mathematics for Economics and Business, Sixth Edition* by Ian Jacques useful for your course. If you are teaching a year-long course, you may wish to recommend both texts. Some adopters have found, however, that they require just one or two extra chapters from one text or would like to select a range of chapters from both texts.

Custom publishing has allowed these adopters to provide access to additional chapters for their students, both online and in print. You can also customise the online resources.

If, once you have had time to review this title, you feel Custom publishing might benefit you and your course, please do get in contact. However minor, or major the change – we can help you out.

For more details on how to make your chapter selection for your course please go to: **www.pearsoned.co.uk/barrow**

You can contact us at: **www.pearsoncustom.co.uk** or via your local representative at: **www.pearsoned.co.uk/replocator**

Introduction

Statistics is a subject which can be (and is) applied to every aspect of our lives. The printed publication *Guide to Official Statistics* is, sadly, no longer produced but the UK Office for National Statistics website[1] categorises data by 'themes', including education, unemployment, social cohesion, maternities and more. Many other agencies, both public and private, national and international, add to this ever-growing volume of data. It seems clear that whatever subject you wish to investigate, there are data available to illuminate your study. However, it is a sad fact that many people do not understand the use of statistics, do not know how to draw proper inferences (conclusions) from them, or misrepresent them. Even (especially?) politicians are not immune from this. As I write the UK referendum campaign on continued EU membership is in full swing, with statistics being used for support rather than illumination. For example, the 'Leave' campaign claims the United Kingdom is more important to the European Union than the EU is to the UK, since the EU exports more to the UK than vice versa. But the correct statistic to use is the *proportion* of exports (relative to GDP). About 45% of UK exports go to the EU but only about 8% of EU exports come to the UK, so the UK is actually the more dependent one. Both sets of figures are factually correct but one side draws the wrong conclusion from them.

People's intuition is often not very good when it comes to statistics – we did not need this ability to evolve, so it is not innate. A majority of people will still believe crime is on the increase even when statistics show unequivocally that it is decreasing. We often take more notice of the single, shocking story than of statistics which count all such events (and find them rare). People also have great difficulty with probability, which is the basis for statistical inference, and hence make erroneous judgements (e.g. how much it is worth investing to improve safety). Once you have studied statistics, you should be less prone to this kind of error.

Two types of statistics

The subject of statistics can usefully be divided into two parts: descriptive statistics (covered in Chapters 1, 10 and 11 of this book) and inferential statistics (Chapters 4–8), which are based upon the theory of probability (Chapters 2 and 3). Descriptive statistics are used to summarise information which would otherwise be too complex to take in, by means of techniques such as averages and graphs. The graph shown in Figure 1.1 is an example, summarising drinking habits in the United Kingdom.

The graph reveals, for instance, that about 43% of men and 57% of women drink between 1 and 10 units of alcohol per week (a unit is roughly equivalent to one glass of wine or half a pint of beer). The graph also shows that men tend to

[1] https://www.ons.gov.uk/

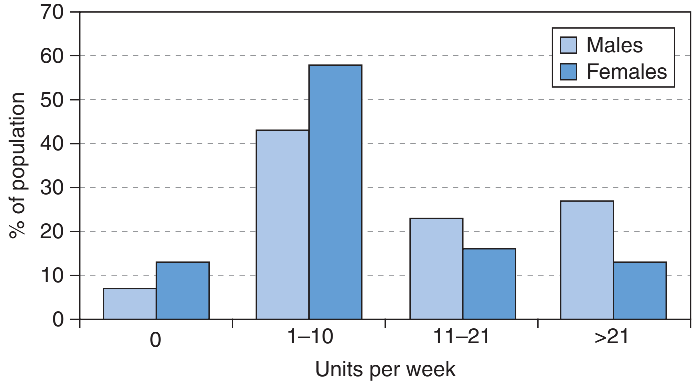


Figure 1.1
Alcohol consumption in the United Kingdom

drink more than women (this is probably no surprise to you), with higher proportions drinking 11 to 20 units and over 21 units per week. This simple graph has summarised a vast amount of information, the consumption levels of about 45 million adults.

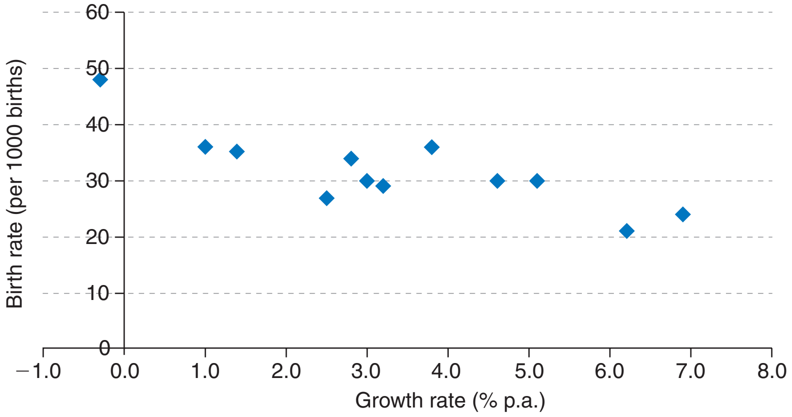
Even so, it is not perfect and much information is hidden. It is not obvious from the graph that the average consumption of men is 16 units per week, of women only 6 units. From the graph, you would probably have expected the averages to be closer together. This shows that graphical and numerical summary measures can complement each other. Graphs can give a very useful visual summary of the information but are not very precise. For example, it is difficult to convey in words the content of a graph; you have to see it. Numerical measures such as the average are more precise and are easier to convey to others. Imagine you had data for student alcohol consumption; how do you think this would compare to the graph? It would be easy to tell someone whether the average is higher or lower, but comparing the graphs is difficult without actually viewing them.

Conversely, the average might not tell you important information. The problem of 'binge' drinking is related not to the average (though it does influence the average) but to extremely high consumption by some individuals. Other numerical measures (or an appropriate graph) are needed to address the issue.

Statistical inference, the second type of statistics covered, concerns the relationship between a sample of data and the population (in the statistical sense, not necessarily human) from which it is drawn. In particular, it asks what inferences can be validly drawn about the population from the sample. Sometimes the sample is not representative of the population (either due to bad sampling procedures or simply due to bad luck) and does not give us a true picture of reality.

The graph above was presented as fact but it is actually based on a sample of individuals, since it would obviously be impossible to ask everyone about their drinking habits. Does it therefore provide a true picture of drinking habits? We can be reasonably confident that it does, for two reasons. First, the government statisticians who collected the data designed the survey carefully, ensuring that all age groups are fairly represented and did not conduct all the interviews in pubs, for example. Second, the sample is a large one (about 10 000 households), so there is little possibility of getting an unrepresentative sample by chance. It would be very unlucky indeed if the sample consisted entirely of teetotallers, for example. We can be reasonably sure, therefore, that the graph is a fair reflection of reality and that the average woman drinks around 6 units of alcohol per week. However,

Figure 1.2
Birth rate v. growth rate



we must remember that there is some uncertainty about this estimate. Statistical inference provides the tools to measure that uncertainty.

The scatter diagram in Figure 1.2 (considered in more detail in Chapter 7) shows the relationship between economic growth and the birth rate in 12 developing countries. It illustrates a negative relationship – higher economic growth appears to be associated with lower birth rates.

Once again we actually have a sample of data, drawn from the population of all countries. What can we infer from the sample? Is it likely that the 'true' relationship (what we would observe if we had all the data) is similar, or do we have an unrepresentative sample? In this case the sample size is quite small and the sampling method is not known, so we might be cautious in our conclusions.

Statistics and you

By the time you have finished this text you will have encountered and, I hope, mastered a range of statistical techniques. However, becoming a competent statistician is about more than learning the techniques, and comes with time and practice. You could go on to learn about the subject at a deeper level and discover some of the many other techniques that are available. However, I believe you can go a long way with the simple methods you learn here, and gain insight into a wide range of problems. A nice quotation relating to this is contained in the article 'Error Correction Models: Specification, Interpretation, Estimation', by G. Alogoskoufis and R. Smith in the *Journal of Economic Surveys*, 1991 (vol. 5, pages 27–128), examining the relationship between wages, prices and other variables. After 19 pages analysing the data using techniques far more advanced than those presented in this book, they state '. . . the range of statistical techniques utilised have not provided us with anything more than we would have got by taking the [. . .] variables and looking at their graphs'. Sometimes advanced techniques are needed, but never underestimate the power of the humble graph.

Beyond a technical mastery of the material, being a statistician encompasses a range of more informal skills which you should endeavour to acquire. I hope that you will learn some of these from reading this text. For example, you should be able to spot errors in analyses presented to you, because your statistical 'intuition' rings a warning bell telling you something is wrong. For example, the *Guardian* newspaper, on its front page, once provided a list of the 'best' schools in England, based on the fact that in each school, every one of its pupils passed a national exam – a 100% success rate. Curiously, all of the schools were relatively small, so perhaps this implies that small schools get better results than large ones? Once you can think statistically you can spot the fallacy in this argument. Try it. The answer is at the end of this introduction.

Here is another example. The UK Department of Health released the following figures about health spending, showing how planned expenditure (in £m) was to increase.

	1998–99	1999–2000	2000–1	2001–2	Total increase over three-year period
Health spending	37 169	40 228	43 129	45 985	17 835

The total increase in the final column seems implausibly large, especially when compared to the level of spending. The increase is about 45% of the level. This should set off the warning bell, once you have a 'feel' for statistics (and, perhaps, a certain degree of cynicism about politics). The 'total increase' is the result of counting the increase from 1998–99 to 1999–2000 *three times*, the increase from 1999–2000 to 2000–1 *twice*, plus the increase from 2000–1 to 2001–2. It therefore measures the *cumulative* extra resources to health care over the whole period, but not the year-on-year increase, which is what many people would interpret it to be.

You will also become aware that data cannot be examined without their context. The context might determine the methods you use to analyse the data, or influence the manner in which the data are collected. For example, the exchange rate and the unemployment rate are two economic variables which behave very differently. The former can change substantially, even on a daily basis, and its movements tend to be unpredictable. Unemployment changes only slowly and if the level is high this month, it is likely to be high again next month. There would be little point in calculating the unemployment rate on a daily basis, yet this makes some sense for the exchange rate. Economic theory tells us quite a lot about these variables even before we begin to look at the data. We should therefore learn to be guided by an appropriate theory when looking at the data – it will usually be a much more effective way to proceed.

Another useful skill is the ability to present and explain statistical concepts and results to others. If you really understand something, you should be able to explain it to someone else – this is often a good test of your own knowledge. Below are two examples of a verbal explanation of the variance (covered in Chapter 1) to illustrate.

Good explanation
The variance of a set of observations expresses how spread out are the data. A low value of the variance indicates that the observations are of similar magnitude, a high value indicates that they are widely spread around the average.

Bad explanation
The variance is a formula for the deviations, which are squared and added up. The differences are from the mean, and divided by n or sometimes by $n - 1$.

The bad explanation is a failed attempt to explain the formula for the variance and gives no insight into what it really is. The good explanation tries to convey the meaning of the variance without worrying about the formula (which is best written down). For a (statistically) unsophisticated audience the explanation is quite useful and might then be supplemented by a few examples.

Statistics can also be written well or badly. Two examples follow, concerning a confidence interval, which is explained in Chapter 4. Do not worry if you do not understand the statistics now.

Good explanation

The 95% confidence interval is given by

$$\bar{x} \pm 1.96 \times \sqrt{s^2/n}$$

Inserting the sample values $\bar{x} = 400$, $s^2 = 1600$ and $n = 30$ into the formula we obtain

$$400 \pm 1.96 \times \sqrt{1600/30}$$

yielding the interval

[385.7, 414.3]

Bad explanation

95% interval $= \bar{x} - 1.96\sqrt{s^2/n} =$

$\bar{x} + 1.96\sqrt{s^2/n} = 0.95$

$= 400 - 1.96\sqrt{1600/30}$ and

$= 400 + 1.96\sqrt{1600/30}$

so we have [385.7, 414.3]

In good statistical writing there is a logical flow to the argument, like a written sentence. It is also concise and precise, without too much extraneous material. The good explanation exhibits these characteristics whereas the bad explanation is simply wrong and incomprehensible, even though the final answer is correct. You should therefore try to note the way the statistical arguments are laid out in this text, as well as take in their content. Chapter 1 contains a short section on how to write good statistical reports.

When you do the exercises at the end of each chapter, try to get another student to read through your work. If they cannot understand the flow or logic of your work, then you have not succeeded in presenting your work sufficiently accurately.

How to use this book

For students:

You will not learn statistics simply by reading through this text. It is more a case of 'learning by doing' and you need to be actively involved by such things as doing the exercises and problems and checking your understanding. There is also material on the website, including further exercises, which you can make use of.

Here is a suggested plan for using the book.

- Take it section by section within each chapter. Do not try to do too much at one sitting.
- First, read the introductory section of the chapter to get an overview of what you are going to learn. Then read through the first section of the chapter trying to follow all the explanation and calculations. Do not be afraid to check the working of the calculations. You can type the data into Excel (it does not take long) to help with calculation.
- Check through the worked example which usually follows. This uses small amounts of data and focuses on the techniques, without repeating all the descriptive explanation. You should be able to follow this fairly easily. If not, work out where you got stuck, then go back and re-read the relevant text. (This is all obvious, in a way, but it's worth saying once.)

- Now have a go at the exercise, to test your understanding. Try to complete the exercise *before* looking at the answer. It is tempting to peek at the answer and convince yourself that you did understand and could have done it correctly. This is not the same as actually doing the exercise – really it is not.
- Next, have a go at the relevant problems at the end of the chapter. Answers to odd-numbered problems are at the back of the book. Your tutor will have answers to the even-numbered problems. Again, if you cannot do a problem, figure out what you are missing and check over it again in the text.
- If you want more practice you can go online and try some of the additional exercises.
- Then, refer back to the learning outcomes to see what you have learnt and what is still left to do.
- Finally – finally – take a deserved break.

Remember – you will probably learn most when you attempt and solve (or fail to) the exercises and problems. That is the critical test. It is also helpful to work with other students rather than only on your own. It is best to attempt the exercises and problems on your own first, but then discuss them with colleagues. If you cannot solve it, someone else probably did. Note also that you can learn a lot from your (and others') mistakes – seeing why a particular answer is wrong is often as informative as getting the right answer.

For lecturers and tutors:

You will obviously choose which chapters to use in your own course, it is not essential to use all of the material. Descriptive statistics material is covered in Chapters 1, 10 and 11; inferential statistics is covered in Chapters 4 to 8, building upon the material on probability in Chapters 2 and 3. Chapter 9 covers sampling methods and might be of interest to some but probably not all.

You can obtain PowerPoint slides to form the basis of you lectures if you wish, and you are free to customize them. The slides contain the main diagrams and charts, plus bullet points of the main features of each chapter.

Students can practise by doing the odd-numbered questions. The even-numbered questions can be set as assignments – the answers are available on request to adopters of the book.

Answer to the 'best' schools problem

A high proportion of small schools appear in the list simply because they are lucky. Consider one school of 20 pupils, another with 1000, where the average ability is similar in both. The large school is highly unlikely to obtain a 100% pass rate, simply because there are so many pupils and (at least) one of them will probably perform badly. With 20 pupils, you have a much better chance of getting them all through. This is just a reflection of the fact that there tends to be greater variability in smaller samples. The schools themselves, and the pupils, are of similar quality.

1 Descriptive statistics

Contents

Contents
continued

Learning outcomes

By the end of this chapter you should be able to:

- recognise different types of data and use appropriate methods to summarise and analyse them
- use graphical techniques to provide a visual summary of one or more data series
- use numerical techniques (such as an average) to summarise data series
- recognise the strengths and limitations of such methods
- recognise the usefulness of data transformations to gain additional insight into a set of data
- be able to write a brief report summarising the data.

Introduction

The aim of descriptive statistical methods is simple: to present information in a clear, concise and accurate manner. The difficulty in analysing many phenomena, be they economic, social or otherwise, is that there is simply too much information for the mind to assimilate. The task of descriptive methods is therefore to summarise all this information and draw out the main features, without distorting the picture.

Consider, for example, the problem of presenting information about the wealth of British citizens (which follows later in this chapter). There are about 18 million adults for whom data are available and to present the data in raw form (i.e. the wealth holdings of each and every person) would be neither useful nor informative (it would take about 30 000 pages of a book, for example). It would be more useful to have much less information, but information which is still representative of the original data. In doing this, much of the original information would be deliberately lost; in fact, descriptive statistics might be described as the art of constructively throwing away much of the data.

There are many ways of summarising data and there are few hard-and-fast rules about how you should proceed. Newspapers and magazines often provide innovative (though not always successful) ways of presenting data. There are, however, a number of techniques which are tried and tested and these are the subject of this chapter. They are successful because: (a) they tell us something useful about the underlying data; and (b) they are reasonably familiar to many people, so we can all talk in a common language. For example, the average tells us about the location of the data and is a familiar concept to most people. For example, young children soon learn to describe their day at school as 'average'.

The appropriate method of analysing the data will depend on a number of factors: the type of data under consideration, the sophistication of the audience and the 'message' which it is intended to convey. One would use different methods to persuade academics of the validity of one's theory about inflation than one would use to persuade consumers that Brand X powder washes whiter than Brand Y. To illustrate the use of the various methods, three different topics are covered in this chapter. First, we look at the relationship between educational attainment and employment prospects. Do higher qualifications improve your employment chances? The data come from people surveyed in 2009, so we have a sample of **cross-section data** giving an illustration of the situation at one point in time. We will look at the distribution of educational attainments amongst those surveyed, as well as the relationship to employment outcomes. In this example, we simply count the numbers of people in different categories (e.g. the number of people with a degree qualification who are employed).

Second, we examine the distribution of wealth in the United Kingdom in 2005. The data are again cross-section, but this time we can use more sophisticated methods since wealth is measured on a **ratio scale**. Someone with £200 000 of wealth is twice as wealthy as someone with £100 000, for example, and there is a meaning to this ratio. In the case of education, one cannot say with any precision that one person is twice as educated as another. The educational categories may be ordered (so one person can be more educated than another, although even that may be ambiguous) but we cannot measure the 'distance' between them. We therefore refer to educational attainment being measured on an **ordinal scale**. In contrast, there is not an obvious natural ordering to the three employment categories (employed, unemployed, inactive), so this is measured on a **nominal scale**.

Third, we look at national spending on investment over the period 1977–2009. This is **time-series data** since we have a number of observations on the variable measured at different points in time. Here it is important to take account of the time dimension of the data: things would look different if the observations were in the order 1977, 1989, 1982, . . . rather than in correct time order. We also look at the relationship between two variables, investment and output, over that period of time and find appropriate methods of presenting it.

In all three cases, we make use of both graphical and numerical methods of summarising the data. Although there are some differences between the methods used in the three cases, these are not watertight compartments: the methods used in one case might also be suitable in another, perhaps with slight modification. Part of the skill of the statistician is to know which methods of analysis and presentation are best suited to each particular problem.

Summarising data using graphical techniques

Education and employment, or, after all this, will you get a job?

We begin by looking at a question which should be of interest to you: how does education affect your chances of getting a job? It is nowadays clear that education improves one's life chances in various ways, one of the possible benefits being that it reduces the chances of being out of work. But by how much does it reduce those chances? We shall use a variety of graphical techniques to explore the question.

The raw data for this investigation come from the *Education and Training Statistics for the UK 2009*. Some of these data are presented in Table 1.1 and show the numbers of people by employment status (either in work, unemployed or inactive, i.e. not seeking work) and by educational qualification (higher education, A levels, other qualification or no qualification). The table gives a **cross-tabulation** of employment status by educational qualification and is simply a count (the **frequency**) of the number of people falling into each of the 12 cells of the table. For example, there were 9 713 000 people in work who had experience of higher education. This is part of a total of nearly 38 million people of working age. Note that the numbers in the table are in thousands, for the sake of clarity.

From the table, we can see some messages from the data; for example, being unemployed or inactive seems to be more prevalent amongst those with lower qualifications: 56% (= (382 + 2112)/4458) of those with no qualifications are unemployed or inactive compared to only about 15% of those with higher education.

However, it is difficult to go through the table by eye and pick out these messages. It is easier to draw some graphs of the data and use them to form conclusions.

The bar chart

The first graphical technique we shall use is the **bar chart**. This is shown in Figure 1.1. The bar chart summarises the educational qualifications of those in work, i.e. the data in the first row of Table 1.1. The four educational categories are arranged along the horizontal (x) axis, while the frequencies are measured

Table 1.1 Economic status and educational qualifications, 2009 (numbers in 000s)

	Higher education	A levels	Other qualification	No qualification	Total
In work	9713	5479	10 173	1965	27 330
Unemployed	394	432	1166	382	2374
Inactive	1256	1440	3277	2112	8085
Total	11 363	7351	14 616	4459	37 789

Source: Adapted from Department for Children, Schools and Families, Education and Training Statistics for the UK 2009, http://dera.ioe.ac.uk/15353/, contains public sector information licensed under the Open Government Licence (OGL) v3.0. http://www.nationalarchives.gov.uk/doc/open-government-licence/open-government

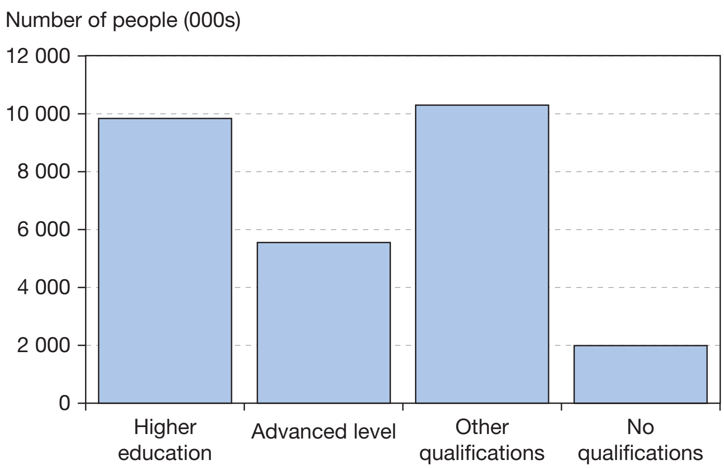
on the vertical (y) axis. The height of each bar represents the numbers in work for that category.

The biggest groups are those with higher education and those with 'other qualifications' which are of approximately equal size. The graph also shows that there are relatively few people working who have no qualifications. It is important to realise what the graph does *not* show: it does not say anything about your likelihood of being in work, given your educational qualifications. For that, we would need to compare the *proportions* of each education category in work; for the moment, we are only looking at the absolute numbers.

It would be interesting to compare the distribution in Figure 1.1 with those for the unemployed and inactive categories. This is done in Figure 1.2, which adds bars for these other two categories.

This **multiple bar chart** shows that, as for the 'in work' category, amongst the inactive and unemployed, the largest group consists of those with 'other' qualifications (which are typically vocational qualifications). These findings simply reflect the fact that 'other qualifications' is the largest category. We can also now begin to see whether more education increases your chance of having a job. For example, compare the height of the 'in work' bar to the 'inactive' bar. It is relatively much higher for those with higher education than for those with

Figure 1.1
Educational qualifications of people in work in the United Kingdom, 2009



Note: The height of each bar is determined by the associated frequency. The first bar is 9713 units high, the second is 5479 units high and so on. The ordering of the bars could be reversed ('no qualifications' becoming the first category) without altering the message.