



© Les Oakshott 2014

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–1–137–34085–6

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

Typeset by MPS Limited, Chennai, India.

CONTENTS

<i>List of Figures</i>	vii
<i>List of Tables</i>	x
<i>Preface to Quantitative Methods</i>	xiii
<i>Acknowledgements</i>	xv
1 COLLECTING DATA: SURVEYS AND SAMPLES	1
2 DESCRIBING DATA USING TABLES AND CHARTS	16
3 DESCRIBING DATA USING SIMPLE STATISTICS	39
4 INDEX NUMBERS	57
5 APPLICATIONS OF PROBABILITY	67
6 PROBABILITY DISTRIBUTIONS	91
7 ANALYSIS OF SAMPLE DATA	108
8 HYPOTHESIS TESTING	122
9 CORRELATION AND REGRESSION	149
10 INVESTMENT APPRAISAL	165
11 TIME SERIES ANALYSIS	178
12 LINEAR PROGRAMMING	197
13 CRITICAL PATH ANALYSIS	210

<i>Appendix 1 Statistical Tables</i>	222
<i>Bibliography</i>	230
<i>Index</i>	231

1

COLLECTING DATA: SURVEYS AND SAMPLES

OBJECTIVES

- To understand the difference between a sampling frame, a sample and a population
- To understand the difference between probabilistic and non-probabilistic sampling
- To know how to select the correct sampling method in different situations
- To understand how to design a questionnaire

INTRODUCTION

Many decisions made by business and by the government are the result of information obtained from sample data, as it is often too costly or impractical to collect data for the whole population.

Data may already exist or it may need to be collected. When we have to collect our own data we call it *primary* data. When it already exists, as in government statistics, we call it *secondary* data. The collection of data can take many forms, but in this chapter we will concentrate on data that is collected by carrying out surveys. During the UK General Election in 2010 many surveys were undertaken by market research companies. The purpose of these surveys was not just to try and understand how people might vote but also to look at attitudes to issues and leaders, for example. Most of these surveys would have been conducted by telephone or face-to-face,

although in some other less urgent surveys people are sent questionnaires by post or email or asked to complete a questionnaire online.

In the surveys undertaken during the UK General Election a lot of care would have gone into selecting a representative sample of the target population. The samples had to be representative of the gender, age profile and ethnicity of the population being surveyed as well as other attributes that might be considered to affect people's voting intentions or views. The number of people being surveyed would also be calculated to ensure that a specified level of accuracy was obtained. A fuller discussion of accuracy issues when taking samples of data is given in Chapter 7.

As well as discussing ways of collecting samples of data this chapter also discusses questionnaire design because it is very important to ensure that the type and wording of questions is given proper attention if the analysis is to be of use.

British Social Attitudes survey

Every year NatCen Social Research (NatCen) undertakes a survey where it asks over 3000 people in Britain what it is like to live and work in the UK. It has been conducted annually since 1983 which makes it NatCen's longest running survey. Respondents are chosen at random using the Post Office Postcode address file. The survey is conducted face to face and is partly paper based and partly by computer. The 2011 survey covered such areas as health, immigration, welfare, Scottish Independence and transport. The interviewer had to complete around an hour's worth of questions and a small sample is given in Table 1.1 below. As well as the questions there are instructions for the interviewer. The sample size for each question is given on the far right of each question (all 3311), and the results of three questions are given as percentages at the end of this table.

Table 1.1 Sample questions from the 2011 British Social Attitudes survey

Q264 [SocSpnd1] * N=3311

CARD B2

Some people think that there should be more government spending on social security, while other people disagree. For each of the groups I read out please say whether you would like to see **more** or **less** government spending on them than now. Bear in mind that if you want more spending, this would probably mean that you would have to pay more taxes. If you want less spending, this would probably mean paying less taxes.

Firstly, ... READ OUT ...

benefits for unemployed people: would you like to see more or less government spending than now?

Q265 [SocSpnd2] *

N=3311

CARD B2 AGAIN

(Would you like to see more or less government spending than now on ...)

... benefits for disabled people who cannot work?

Q266 [SocSpnd3] *

N=3311

CARD B2 AGAIN

(Would you like to see more or less government spending than now on ...)

... benefits for parents who work on very low incomes?

Q267 [SocSpnd4] *

N=3311

CARD B2 AGAIN

(Would you like to see more or less government spending than now on ...)

... benefits for single parents?

Q268 [SocSpnd5] *

N=3311

CARD B2 AGAIN

(Would you like to see more or less government spending than now on ...)

.. benefits for retired people?

Q269 [SocSpnd6] *

N=3311

CARD B2 AGAIN

(Would you like to see more or less government spending than now on ...)

... benefits for people who care for those who are sick or disabled?

(continued)

Table 1.1 Continued

	SocSpnd1	SocSpnd2	SocSpnd3
	%	%	%
Spend much more	1.2	5.6	5.1
Spend more	13.8	47.1	53.2
Spend same as now	32.3	39.3	34.3
Spend less	40.9	5.0	4.5
Spend much less	9.7	0.4	0.5
Don't know	2.1	2.6	2.3
(Refusal)	0.1	0.1	0.0

Source: NatCen Social Research's British Social Attitudes survey

THE BASICS OF SAMPLING

A survey only collects information about a small subset of the *population*. The word 'population' can and often does refer to all the people in Britain or a town, but for statisticians it is also a general term used to refer to all groups or items being surveyed. For instance, it could refer to the viewing habits of all children in a town or, as you will see in a later chapter, it could refer to the weights of jars of coffee produced by a company during a week. The alternative to a survey is to question every member of the population, and when this is done it is called a *census*. Unfortunately it is expensive and very difficult to carry out a census, and also unnecessary. A survey of a small subset of the population, called a *sample*, can give surprisingly accurate results if carried out properly. This and other chapters will show you what factors must be taken into account to give accurate results.

It is crucial to be clear about the purpose of the survey as this will define your target population. Once you have selected your target population, you need to determine whether there is any list that would allow you to identify every member of the population. This list is called a *sampling frame*, and examples include the electoral register, the postcode address file, a company's personnel records or even a list of all serial numbers of cars built by one car manufacturer last year. Sometimes a sampling frame is simply not

available or is too difficult to obtain, in which case achieving a representative sample will be more difficult but not necessarily impossible.

Once your target population has been chosen and an appropriate sampling frame identified, it is necessary to choose your sample. If the sample is chosen badly your results will be inaccurate due to *bias* in your sample. Bias is caused by choosing a sample that is unrepresentative of the target population. To avoid bias you need to ensure that your sample is representative of the target population. You will see how this can be achieved later.

The purpose of a survey is to obtain information about a population. All other things being equal, the accuracy of the sample results will depend on the sample size; the larger the sample, the more accurate the results. A large sample will clearly cost more than a small one, although the method that is employed to collect the data will also determine the accuracy and cost of the survey. Methods of data collection range from the use of postal questionnaires to 'face-to-face' interviews. Some methods of data collection are expensive but guarantee a good response rate, while others are cheap to administer but are likely to produce quite a poor response.

Questionnaire design

Questionnaire design is more of an art than a science and there is no universal design that would be suitable for all situations. The actual design will depend on factors such as:

- The type of respondent (for example, business, consumers, children)
- The method of data collection (postal, telephone, face-to-face or online)
- The resources available

Questions are of two forms: close-ended questions and open-ended questions. Close-ended questions give the respondents a choice of answers and are generally considered much easier to answer and to analyse. However, the limited response range can give misleading results. Open-ended questions, such as 'Why did you buy this product?' allow respondents more flexibility in the type of response (you may get answers that you hadn't thought of),

method of randomly selecting the required sample from this list. The simplest example of this technique is a raffle where the winning ticket is drawn from the ‘hat’. For a more formal application a stream of random numbers would be used. Random numbers are numbers that show no pattern; each digit is equally likely. A table of random numbers is given in Appendix 1 or you can use the RAND function in Excel. The method of simple random sampling using random numbers is quite easy to apply, although tedious, as you will see from Example 1.1.

Example 1.1

Table 1.2 is a part of a list of students enrolled on a business studies course at a university.

Table 1.2 List of students enrolled on a business studies course

Number	Name	Gender
1	N. Adams	Male
2	C. Shah	Male
3	B. Booth	Female
4	C. Meng	Male
5	A. Ho	Male
6	D. Drew	Male
7	K. Fisher	Female
8	P. Frome	Male
9	G. Godfrey	Male
10	J. Bakoulas	Male
11	D. Jeffrey	Female
12	H. Jones	Male
13	M. Li	Male
14	N. King	Female
15	K. Lenow	Male
16	A. Loft	Female
17	T. Georgiou	Female
18	S. Moore	Female
19	F. Muper	Female
20	R. Muster	Female
21	A. Night	Male
22	J. Nott	Male
23	L. Nupper	Male
24	K. Khan	Male
25	O. Patter	Female

Say from this 'population' of 25 students you wanted to randomly select a sample of 5 students. How would you do it? You could use the student number (in this case conveniently numbered from 1 to 25) and then try and obtain a match using a stream of two-digit random numbers between 0 and 25. For example, suppose you had the following random numbers: 10, 25, 8, 23, 15, 21, 12, 19, 4, 15, 7. The first five numbers equate to J. Bakoulas, O. Patter, P. Frome, L. Nupper and K. Lenow.

One question that we should ask of any sample is how representative it is of the target population. In our population we have 15 males and 10 females which is a proportion of 60% males to 40% females. In the sample we have 4 males out of 5 which is a proportion of 80%, and on this statistic our sample doesn't represent the target population very well. Another sample might give us a completely different proportion and you could even get a sample of all the same gender. This variation is called sampling error and occurs in all sampling procedures. In Chapter 7 you will be shown how to control and quantify this error.

Stratified sampling

Many populations can be divided into different categories. For example, a population of adults consists of the two sexes, employment status, ethnicity and many other categories. If you think that the responses you will get from your survey are likely to be determined partly by each category, then clearly you want your sample to contain each category in the correct proportions.

You probably realized that the 'perfect' sample above should contain 3 males (60% of 5). In order to ensure that you will get exactly 3 males, you should first of all have separated out the two sexes and then obtained two simple random samples, one of size 3 and one of size 2, as shown in Table 1.3. The two populations have been re-numbered, although this is not essential. Using the same random numbers as before and ignoring numbers greater than 15 for the male sample and 10 for the female sample, we get male students 10, 8 and 15; that is M. Li, K. Khan and C. Shah, and female students 4 and 7, that is; D. Jeffrey and S. Moore.

Table 1.3 Table ordered by gender**Male students**

Number	Name
1	N. Adams
2	J. Bakoulas
3	D. Drew
4	P. Frome
5	G. Godfrey
6	A. Ho
7	H. Jones
8	K. Khan
9	K. Lenow
10	M. Li
11	C. Meng
12	A. Night
13	J. Nott
14	L. Nupper
15	C. Shah

Female students

Number	Name
1	B. Booth
2	K. Fisher
3	T. Georgiou
4	D. Jeffrey
5	N. King
6	A. Loft
7	S. Moore
8	F. Muper
9	R. Muster
10	O. Patter

Stratified sampling is a very reliable method, but it does assume that you have a knowledge of the categories of the population. Stratified sampling is often used in conjunction with the next method.

Multi-stage sampling

If the target population covers a wide geographical area then a simple random sample may have selected respondents in quite different parts of the country. If the method employed to collect the data is of the face-to-face interview type, then clearly a great deal of travelling could be involved. To overcome this problem the area to be surveyed is divided into smaller areas and a number of these smaller areas randomly selected. If desired, the smaller areas chosen could themselves be divided into smaller

districts and a random number of these selected. This procedure is continued until the area is small enough for a simple random sample (or a stratified sample) to be selected. The final sample should consist of respondents concentrated into a small number of areas. It is important that the random sample chosen from each area is the same proportion of the population or bias towards certain areas could result. As it is, bias is likely to occur as a result of similarity of responses from people within the same area, but this is the price you pay for reduced travelling time.

Example 1.2

A broadcasting company wishes to obtain a representative sample of television viewers from across Britain using multi-stage sampling.

The country could be split into counties, or perhaps television regions might be more appropriate in this case. A number of these would be chosen at random, and these areas subdivided into district councils. A random sample of districts within each chosen region could now be selected, and the selected districts divided into postal areas. A random sample of residents within each chosen postal area could then be chosen using the register of electors.

Figure 1.1 illustrates this process in diagrammatic form. At each level you are taking a random sample. Note that it is not until you get to the individual elector that you carry out the actual survey.

Cluster sampling

Cluster sampling is similar to multi-stage sampling and is used when a sampling frame is not available. Again a large geographical area is divided into a number of smaller areas called clusters. If necessary these clusters can be further subdivided to obtain clusters which are small enough for all members of the cluster to be surveyed. As with multi-stage sampling, a bias will result due to similarities in responses from members of the same cluster. The difference between cluster sampling and multi-stage sampling is that since individual members of a cluster cannot be identified in advance, it is necessary for all members to be surveyed. Random sampling is therefore not involved.

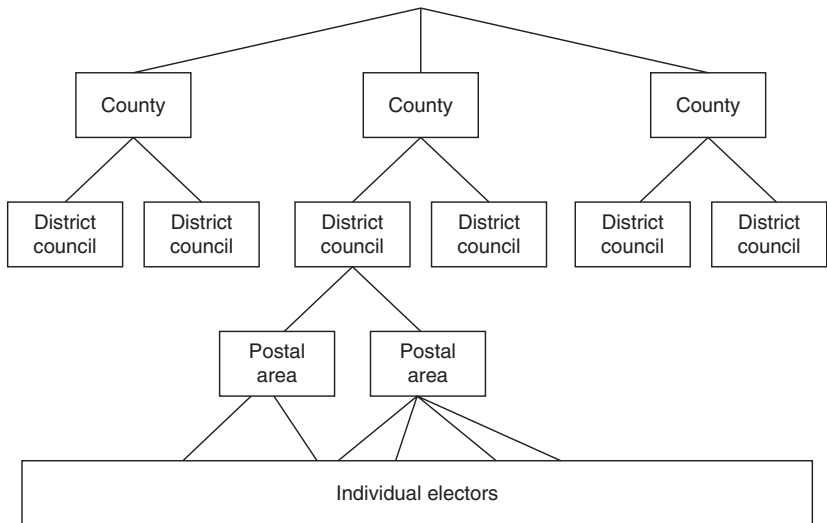


Figure 1.1 Multi-stage sampling

Systematic sampling

This method is normally used with a sampling frame, but it can also be used where a sampling frame is not available. However, in this case the size of the population must be known. The idea is that every n th member of a population is selected, where the value of n is determined by the size of the population and by the required sample size. For instance, if a 5% sample is to be selected from a population of size 1000, then every 50th person will be selected. The start of the sequence is usually chosen at random. For example, if a 20% sample was to be selected from the student population given in Table 1.2 every 5th person would be selected. If you started with, say, the third student, your sample would consist of B. Booth, P. Frome, M. Li, S. Moore and L. Nupper.

Systematic sampling is a very quick and efficient method of obtaining a sample. The sample should be random, provided there is no pattern in the way people are ordered in the population.

Quota sampling

I am sure that you have seen an interviewer in a town centre with a clipboard waiting to pounce on some unsuspecting individual! The interviewer is in fact looking for particular groups of individuals who meet the categories that he or she has been asked to interview. Within each group there will be a number or quota of people required, and the survey is complete when the quotas have been reached. Quota sampling is a non-probabilistic version of stratified sampling. The quotas within each group should, like stratified sampling, reflect the proportions within the target population.

Other sampling methods

There are three other non-probabilistic sampling methods that are sometimes used. These are judgemental, purposive and snowball sampling. With judgemental sampling the researcher makes a judgement about what constitutes a representative sample. If a government agency was interested in the effects on people's health of car exhaust fumes, they would choose areas near cities or motorways to obtain the sample. They would not choose rural areas, except perhaps for a control group. Purposive sampling is where certain members of the population are purposefully chosen. For example, customers holding store loyalty cards might be asked about planned improvements to the store. Snowballing is where a sample is chosen using one of the methods mentioned in this chapter and then additional members of the population are generated from this sample. An example could be in the investigation of the mis-selling of pensions that occurred in Britain during the late 1980s. A sample of pensioners could be obtained and any person who was persuaded to leave their occupational pension scheme would be asked to name other people they knew who were also affected. In this way the sample size could be increased.

KEY POINTS

It is generally impractical to question or observe every member of the target population, so a sample of this population is selected instead.

Table 1.4 A summary of the sampling methods available

	Sampling frame available (probabilistic sampling)	Sampling frame not available (non-probabilistic sampling)
Population resides in one place	Simple random sampling Systematic sampling	Systematic sampling (if the size of the population is known) Judgemental sampling Purposive sampling
Population geographically scattered	Multi-stage sampling	Cluster sampling Judgemental sampling
Population is defined by categories	Stratified sampling	Quota sampling
Population is small and unknown		Snowballing

To obtain data from a population of people you normally carry out a survey. This survey can be done by post, by telephone, face-to-face and online. Each method has its advantages and disadvantages. A face-to-face survey normally gives you the best response rate but is more costly than other methods.

Whatever method you use the sample should be representative of the target population.

- ★ Probabilistic sampling methods will give you a representative sample but require a sampling frame.
- ★ Non-probabilistic sampling is generally quicker to carry out but may not be completely representative of the target population.

Table 1.4 summarizes the different sampling methods available.

FURTHER READING

There are a huge number of quantitative methods and statistical texts that cover the collection of data. Some like Morris (2008), Curwin and Slater (2008) and Chapter 3 in Oakshott (2012) cover similar material to this

text. Collis and Hussey (2009) is a more general text on business research methods. Fowler (2009) is a very detailed text on survey research methods and would be ideal for anyone having to design their own survey.

REVISION QUESTIONS

- 1 A town with historical connections has received a grant of £20m in order to improve its tourist facilities. The town councillors have decided to ask a representative sample of residents how the money should be spent. Given that expenditure for the survey must be kept to a minimum suggests ways in which a representative sample of residents can be chosen.
- 2 Which of the following are likely to have a sampling frame?
 - Students at a university
 - Employees of a company
 - Concert goers
 - Shoppers at a shopping mall
 - Members of a social networking group
- 3 A company wishes to carry out a survey of its employees to monitor their views on the future of the company. A departmental breakdown of the company's 200 employees is as follows:

Shop-floor/warehouse	80
Service engineers	15
Quality control	20
Marketing and sales	25
Accounts	15
Personnel	10
Administration	25
Catering	10

A survey of 40 employees is to be conducted. A sampling frame is available, listing the employees by surname in alphabetical order, independent of department.

- (a) Explain how the following sampling methods could be carried out to obtain the sample of 40 employees:
- simple random sampling
 - systematic sampling
 - stratified sampling
 - quota sampling
- (b) Discuss the benefits and drawbacks of using quota sampling to obtain the sample of 40 employees.
- 4** You have been asked to conduct a survey into the attitudes of school leavers to higher education. You intend to carry this out using the face-to-face interview method. How would you obtain your sample?

INDEX

- χ^2 distribution *see*
 - chi-square distribution
- χ^2 hypothesis test *see*
 - chi-square hypothesis test
- a priori* approach to
 - probability 70
- accounting rate of return
 - 170
- activity on node method
 - 212–16
- addition rule of probability
 - 72
- additive time series model
 - 185–6
- aggregate index *see*
 - weighted aggregate index
- ARR *see* accounting rate of return
- Averages 41–8
 - Mean 42–8
 - Median 43–7
 - Mode 43, 45, 47
- bar charts 28–31
 - component 30
 - multiple 30
 - percentage 31
 - simple 29
- Bayes' theorem 77–9, 83–4
- binomial distribution 95–8
 - approximating by the normal distribution 104–5
 - approximating by the Poisson distribution 99
 - mean and standard deviation of 98
- box and whisker plot 54
- categorical data 19
- central limit theorem 111
- central tendency, measures of *see* averages
- centred moving averages
 - 183
- charts
 - bar 28–31
 - box and whisker 54
 - frequency polygon 32
 - line graph 32
 - ogive 34
 - pie 27–8
 - scatter diagrams 151–5
 - stem and leaf plot 34–6
 - tally 24
- chi square distribution
 - 140–1
- chi-square hypothesis test
 - 140–5
 - goodness-of-fit test 141–3
 - test of association 143–5
- cluster sampling 10
- coefficient of correlation
 - see* correlation
- coefficient of determination
 - 161
- coefficient of variation 53
- collection of data 1–12
- component bar chart 30
- compound events probability of 70–9
- compound interest 171
- conditional probability
 - 74–9
- confidence intervals
 - 112–19
 - of a mean 112–16
 - of a percentage 117–18
- consumer price index 63–4
- continuous data 19
- continuous probability distribution 93, 100
- correlation 151–8
 - Pearson's product moment correlation 157–8
 - scatter diagrams 151–5
 - Spearman's rank correlation coefficient 156–7
- CPA *see* critical path analysis
- crashing 216–18
- critical path 216
- critical path analysis
 - 210–18
 - activity-on-node 212–16
 - cost scheduling 216–18
 - crashing 216–18
 - critical path 216
 - float 212
- critical region 124, 128, 140
- cumulative frequency ogive
 - see* ogive

- data
 - categorical 19
 - collection of 1–12
 - continuous 19
 - discrete 19
 - grouped 22–6
 - interval scale 19, 157
 - nominal 19
 - ordinal 19
 - outlier 36
 - presentation of 16–37
 - primary 1
 - ratio scale 19
 - secondary 1
 - summarising 39–54
 - tabulation 20–6
- decision trees 80–3
 - and Bayes' theorem 83–6
- diagrammatic
 - representation of data 26–36
- discounted cash flow
 - techniques 171–5
- discrete data 19
- discrete probability
 - distributions 93–9
- dispersion *see* spread, measures of

- empirical probability 70
- EMV *see* expected monetary value
- expected monetary value 80
- expected value 79, 141, 143
- exponential smoothing 191–3

- forecasting *see* Time series analysis
- frame, sampling 4, 6, 10, 11, 13

- frequency distributions
 - empirical 94
 - mean of 44, 46
 - median of 44, 46–7
 - mode of 45, 47
 - standard deviation 52–3
- frequency polygon 32
- goodness-of-fit test 141–3
- histogram 32–3
- hypothesis tests 122–46
 - chi-square hypothesis tests 140–5
 - goodness-of-fit test 141–3
 - large sample test for a population mean 125–7
 - large sample test for two independent population means 132–3
 - paired t-test 136–8
 - population percentage 130–1
 - small sample test for a population mean 127–9
 - small sample test for two independent population means 133–6
 - test of association 143–5
 - two population percentages 138–40
- independent events 71
- index numbers 57–64
 - consumer price index 63–4
 - Laspeyres' index 62
 - Paasche's index 62–3
 - retail price index 63–4
 - simple indices 59–60
 - weighted aggregate 60–3
- inference *see* confidence intervals and hypothesis tests
- internal rate of return 174–5
- interquartile range 49
- interval estimates *see* confidence intervals
- interval scale data 19, 157
- investment appraisal 165–76
 - accounting rate of return 170
 - compound interest 171
 - discounted cash flow techniques 171–5
 - internal rate of return 174–5
 - net present value 172–3
 - payback method 170
 - present value 172
 - simple interest 171
- IRR *see* internal rate of return

- judgemental sampling 12

- large sample test for a population mean 125–7
- large sample test for two independent population means 132–3
- Laspeyres' index 62
- least squares method 160–1
- line graph 32
- linear programming 197–207

- binding and non-binding constraints 203
- feasible region 203
- graphical solution 201–3
- model formulation 199–201
- tight and slack constraints 203–4
- linear regression 158–62
 - coefficient of determination 161–2
 - least squares method 160–1
- location, measures of *see* averages
- LP *see* linear programming
- Mean 42–8
 - of a binomial distribution 98
 - of a frequency distribution 44, 46
 - of a normal distribution 100
 - of a Poisson distribution 98
 - weighted 48
- mean square error 189
- median 43–7
- mode 43, 45, 47
- moving averages 183
- multiple bar chart 30
- multiplication rule of probability 74
- multiplicative time series model 186–7
- multi-stage sampling 9–10
- net present value 172–3
- network analysis *see* critical path analysis
- nominal data 19
- non-probabilistic sampling 12
- normal distribution 100–5
 - approximating a binomial distribution 104–5
 - approximating a Poisson distribution 104–5
 - standard normal distribution 101–2
- NPV *see* net present value
- ogive 34
- ordinal data 19
- outlier 36
- Paasche's index 62–3
- paired t-test 136–8
- payback method of investment appraisal 170
- Pearson's product moment correlation coefficient 157–8
- percentage bar chart 31
- pie chart 27–8
- point estimates 110
- Poisson distribution 98–100
 - mean and standard deviation of 100
- posterior probability 77
- present value 172
- presentation of data 16–37
 - diagrammatic methods 26–36
 - tabulation of 20–6
- primary data 1
- prior probability 77, 78, 83, 84
- probabilistic sampling 12, 109
- probability 67–87
 - a priori 70
 - addition rule 72
 - Bayes' theorem 77–9
 - compound events 70–9
 - conditional 74–9
 - distributions 91–106
 - empirical 70
 - expected value 79
 - independent events 71
 - multiplication rule 74
 - mutually exclusive events 71
 - mutually exhaustive outcomes 71
 - posterior 77
 - prior 77, 78, 83, 84
 - subjective 69
 - tree diagrams 75–77, 78
 - Venn diagrams 72–3
- probability distributions 91–105
 - binomial 95–8
 - chi-square 140–1
 - continuous 93, 100
 - discrete 93–9
 - empirical 94
 - normal 100–5
 - Poisson 98–100
- purposive sampling 12
- questionnaire design 5–6
- quota sampling 12
- random numbers 7
- range 49
- ratio scale data 19
- regression *see* linear regression
- retail price index 63–4
- return on capital employed *see* accounting rate of return
- sampling 4–12
- sampling distribution of the mean 110
- of a percentage 117
- sampling error 8

- sampling frame 4
- sampling methods 6–12
 - cluster 10
 - judgemental 12
 - multi-stage 9–10
 - non-probabilistic 12
 - probabilistic 6–10
 - purposive 12
 - quota 12
 - simple random 6–8
 - snowball 12
 - stratified 8–9
 - systematic 11
- scatter diagrams 151–5
- secondary data 1
- significance testing *see* hypothesis testing
- simple bar chart 29
- simple indices 59–60
- simple interest 171
- simple random sampling 6–8
- small sample test for a population mean 127–9
- small sample test for two independent population means 133–6
- snowball sampling 12
- Spearman's rank correlation coefficient 156–7
- spread, measures of 48–54
 - interquartile range 49
 - range 49
- standard deviation 49–53
 - variance 49–53
- standard deviation 49–53
 - of a binomial distribution 98
 - of a frequency distribution 52–3
 - of a normal distribution 100
 - of a Poisson distribution 100
 - of a population 52, 110, 125, 132, 134
 - of a sample 49, 110, 111, 114, 128, 132
- standard error
 - of paired sample 136
 - of a percentage 117, 130
 - of the mean 111, 126, 185
 - of two independent means 132, 134
 - of two population percentages 138
- standard normal distribution 101–2
- stem and leaf plot 34–6
- stratified sampling 8–9
- subjective probability 69
- summarising data 39–54
 - averages 41–8
 - spread 48–53
- survey methods *see* sampling methods
- systematic sampling 11
- t-distribution 115
- t-test for a mean 127–9
- tabulation of data 20–6
- tally chart 24
- test of association 143–5
- tests of hypothesis *see* hypothesis tests
- time series analysis 178–94
 - additive model 185–6
 - analysis of errors 187–9
 - centred moving averages 183
 - exponential smoothing 191–3
 - forecasting 189–90, 191–2
 - mean square error 189
 - moving averages 183
 - multiplicative model 186–7
 - seasonal component 184–7
 - trend 183–4
- tree diagrams 75–7, 78
- variance 49–53
- Venn diagram 72–3
- weighted aggregate index 60–3
- weighted mean 48
- Z values 101–4
- Z test *see* hypothesis test