STATISTICS

ADVANCED MATHEMATI

LONGMAN

Nuffield Advanced Mathematics

Statistics

LONGMAN

Science Learning Centres
N12617

Acknowledgements

The production of this series of advanced mathematics books was made possible by the following grants and support.

1 Generous grants from The Nuffield Foundation and from Sir John Cass's Foundation.

2 Valuable support from

- Northumberland County Council, for seconding Bob Summers to work with the project, and the unstinting help and encouragement of Chris Boothroyd, Director of their Supported Self Study Unit
- David Johnson, Licentiate Professor, Centre for Educational Studies, King's College, London
- the Local Education Authorities of Camden, Essex, Islington, Northumberland, Oxford, Tower Hamlets and Westminster for releasing teachers to work with the project
- Afzal Ahmed, Warwick Evans, Adrian Oldknow and Honor Williams from the Mathematics Centre, West Sussex Institute of Higher Education
- Geoff Wake from the Mechanics in Action Project, University of Manchester.

This course has been written with the help of many people in addition to the team, which consists of: Hugh Neill, Director; Sue Burns, Deputy Director; Sue Ahrens and Bob Summers, Development Officers; and Nina Towndrow, Project Administrator.

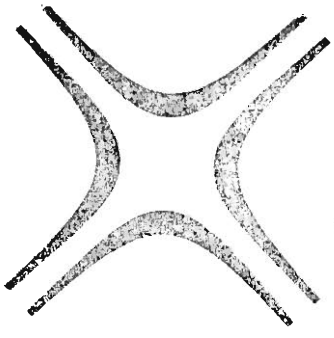
Other contributing authors are: Mary Barnes, Jill Bruce, Paul Garcia, Adrian Oldknow, Tony Ralston, Mary Rouncefield, Geoff Wake and Peter Wilder.

Teachers who have also helped in the production of materials include: Paul Aljabar, Jane Annets, Dinny Barker, Richard Choat, Joel and Trudy Cunningham, Jan Dangerfield, John Deakin, Phil Donovan, Jenny Douglas, John Eyles, Jan Galloway, Peter Horril, Paul Jenkins, Sue Maunder and Mike Warne.

The general editor was Hugh Neill.

We are grateful to Polity Press, Catherine Marsh, *Exploring Data – An Introduction to Data Analysis for Social Scientists*, 1988 for permission to reproduce Appendix A.

The photograph on page 22 was taken by John Birdsall.



NATIONAL
STEM
CENTRE

Longman Group UK Limited
Longman House, Burnt Mill, Harlow, Essex CM20 2JE,
England and Associated Companies throughout the World.

© Nuffield Foundation 1994

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the prior written permission of the Publishers or a licence permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, W1P 9HE.

First published 1994
ISBN 0582 09985 4

Set in 10/12 Times by Stephen I Pargeter
Illustrated by Oxford Illustrators Limited
Produced by Longman Singapore Publishers Pte Ltd
Printed in Singapore

The publisher's policy is to use paper manufactured from sustainable forests.

Contents

Statistics

1	Introduction	*1*
2	Collecting real data	*7*
3	Looking for differences	*33*
4	Fitting a model to the data	*54*
5	The Poisson distribution	*83*
6	Simulations	*95*
7	Summaries and exercises	*100*
8	Hints	*105*
9	Answers	*112*
	Appendix A The HEIGHT dataset	*138*
	Appendix B Tables	*141*
	Index	*148*

Using Nuffield Advanced Mathematics

- The book consists mainly of activities through which you can develop your understanding of mathematical ideas and results, or apply your knowledge and understanding to problems of various kinds.
- The activities are designed so that you can use your graphics calculator fully, gaining all the advantages that this can bring to your learning of mathematics.
- For most of the activities you can work either individually or in a small group of two to four students, but there are some activities for which individual or group work is specifically recommended.
- You will be working on the activities outside the classroom as well as in class, sometimes with and sometimes without help from your teacher.
- Do not assume that you should work through every activity in a unit. The Nuffield course is suitable for a variety of students with a variety of mathematical backgrounds and interests.
- **Planning your work** on each unit is essential, as you will need to decide together with your teacher, which activities to work on, and whether to work on them in class or as part of your independent study time.
- The information in unit and chapter introductions will help with your planning. The **chapter summaries** and **practice exercises** at the end of the book will also be useful for this purpose. For example, you might already know all of the results that are developed in some particular chapter, in which case it might be appropriate to skip that chapter.
- On several occasions during your study of a unit, you will take part in a **review**. Reviews can take a number of forms: their main purpose is for you to take stock, through discussion with your teacher and other students, of what you have learned so far.
- To assist with the review process, the end of every chapter contains some specific suggestions which might form the basis for discussion; there is also a list headed **What you should know**. This list is a reminder of the main ideas of the chapter and of any new mathematics terms that it has introduced. New mathematical terms are written in **bold** type the first time they are explained in the unit. Most chapters end with some check questions for you to tackle to make sure you have grasped the main ideas of the chapter. There are no answers provided for these questions, but answers are provided for nearly all the other questions on which you will work.
- You will find messages at various stages in the materials suggesting for example that you consult your teacher, or should retain the results of a particular activity for later use. **These will be printed in a different typeface so that you can recognise them.** In addition, there will be general mathematical comments which appear in shaded boxes. These are usually to give you extra background information.
- Towards the back of the book there is a chapter of hints for various activities if you need help in getting started. You will find the hint symbol  in the margin next to the particular question for which there is a hint. You should only look up the hint if you find you are unable to solve the question without it.

1 Introduction

In earlier units about statistics you have collected data, explored datasets, set up models, analysed them and made predictions. In the Investigating data unit in Book 2 you used exploratory methods to analyse data; from your analysis you were able to suggest possible hypotheses. In the Correlation and prediction unit in Book 4 you learned how to model the relationship between two variables with straight lines. The Distributions unit in Book 5 introduced you to some probability distributions for modelling data, and showed you the ideas of testing hypotheses and estimating by using confidence intervals. In this option you will develop some of these ideas, in preparation for conducting a statistical investigation of your own choice.

In earlier units you have been used to thinking of the modelling process as one in which you move back and forth between a context and a mathematical model. This process has been represented by the following cyclical diagram.

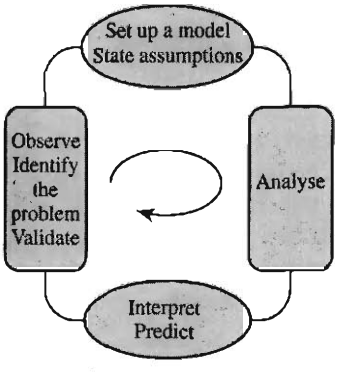


Figure 1.1

The first time round the modelling cycle is often exploratory. The researcher makes observations, and begins to collect information to see whether it fits a pattern. For example, in the 1940s research began into the increasing number of cases of lung cancer. The focus at that stage was simply on collecting data about cases of lung cancer, but the early evidence showed a clear link with tobacco smoking, although it was still not certain that one caused the other. Thus, the interpretation of the early results led to a tentative hypothesis that smoking was associated with lung cancer, and researchers began to collect more figures to see if they fitted the same pattern.

Later investigations looked first for evidence that smoking caused lung cancer, and second for associations between smoking and other diseases. Tobacco smoking became widespread during and after the First World War, so it was not until the 1950s that there was evidence from a large number of people to support a more concrete hypothesis about the effects of long-term smoking. A newspaper report (the *Guardian*, 16 February 1993) claims that today the evidence is that smokers are three times more likely than non-smokers to die between the ages of 55 and 64, and that tobacco smoking causes 30 times as many deaths as traffic accidents.

Activity 1.1 Smoking evidence

1 Consider the early investigation of the effects of smoking, described above. Identify each stage of the statistical process as the investigation unfolded.

2 Read again the two claims made at the end of the paragraph just before this activity. What extra information would you need about the evidence behind these

two claims, and how it was collected, before you could understand what the claims really mean?

The structure of this book

You can think of the process of statistical investigation as a cyclical process with four stages

- define the problem
- collect the data
- analyse the data
- interpret the results.

The first four chapters of this book are grouped according to the four stages of statistical investigation.

In Chapter 1 you will find some examples of problems investigated by statisticians, which show some of the issues involved in defining and refining the problem so that the investigation can produce an answer.

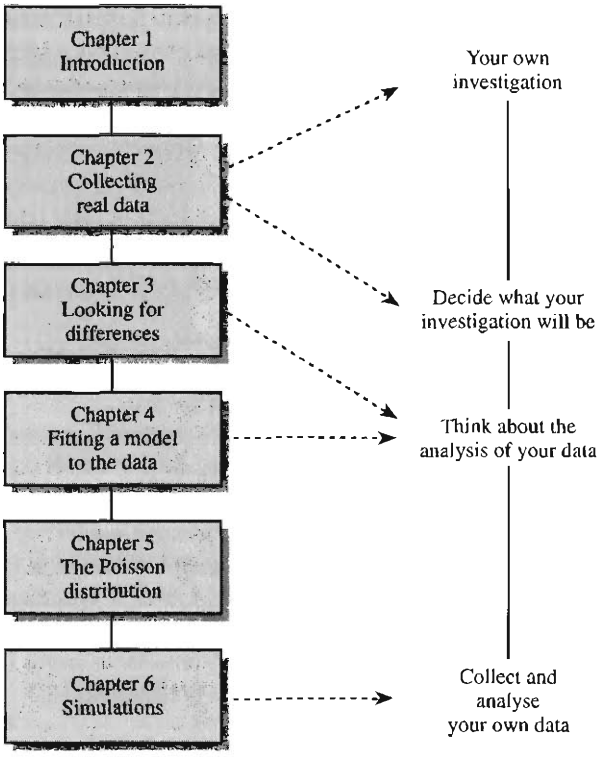
In Chapter 2 you will learn about ways of collecting data to tackle the problem. There are many decisions to be made at this stage.

- Do you need to carry out a survey or conduct a controlled experiment?
- Are you going to survey the whole population or can you take a sample? If so, how should you choose the sample?
- How will you measure the variables that you want to measure?
- What factors, other than those in which you are interested, affect the data? Can you control any of them?
- Do you need a questionnaire?

While you are working on Chapter 2, begin thinking about a problem which you would like to investigate for yourself, and make some preliminary decisions about how you will collect your data. Then, as you work through the rest of the book, you will be able to consider which techniques for analysing data will be appropriate for your investigation. The final section in Chapter 2 is about questionnaires: study this carefully if you choose an investigation which needs a questionnaire.

In Chapters 3 and 4 you will learn how to analyse data and interpret the results. In Chapter 3 you will find methods for testing for differences between two samples. In Chapter 4 you will find some methods for testing how well a particular theoretical model fits the observed data.

The final two chapters are not essential reading, but you may need to work through them in preparation for your own investigation. Chapter 5 introduces another distribution, building on the work that you did in the *Distributions* unit in Book 5. In Chapter 6 you will find an introduction to the idea of statistical simulation as a way to predict what will happen if changes are made in an experiment or other situation.



The chapters are not all the same length. Chapter 1 is short and should take no more than five hours. You need to spend part of this time following up some of the suggestions made later in the chapter for your own investigation.

Chapter 2 is much longer and might take up to 25 hours of your time. At the end of Chapter 2 you need to spend some time planning the data collection for your own investigation.

Chapter 3 will need between 15 and 20 hours, and Chapter 4 may need a little longer.

Chapter 5 will also need about 15 hours, and Chapter 6 about 10 hours, but these are optional. Your own investigation should represent an additional 20 hours work, including data collection, analysis and writing a report.

All timings are for your learning time, approximately half of which will take place out of the classroom.

There are summaries and further practice exercises for chapters 2, 3, 4 and 5 at the back of this book (Chapter 7).

Your own investigation

At the end of Chapter 2, in Activity 2.20, you will be asked to choose a topic for your own investigation. You should discuss your choice with your teacher in your review at the end of Chapter 2, and consider what methods of data collection you will need to use.

To prepare yourself to make your choice of topic, you need to begin considering this matter now. Some suggestions of possible topics are given below. Many of the suggestions involve complex situations, so you will need to be precise about the question you choose to research. Remember that your time and resources are limited, and you will need to complete your research and report on your findings before the end of the course.

As you work through the rest of this chapter, and Chapter 2, you should keep your own investigation in mind. Many of the following suggestions are referred to in activities in later chapters, and where this is so the reference is given.

Suggestions

1 Investigate the ways in which people discriminate between the sexes. Activity 2.19 might be a starting point for this.

2 Investigate young people's awareness of health issues. You will do some work on this in Activities 2.16 and 2.17.

3 Do people weigh more last thing at night than they do first thing in the morning? (See Activities 2.8, question **3** and 3.3, question **1**.)

4 Investigate whether it is easier to recognise that two letters of the alphabet are the same if they are written in the same typeface than if they are written in a different typeface. (See Activity 2.8, question **2**.)

5 Do people find it easier to remember words or facts which make sense to them, as opposed to those which do not? Activities 3.1 and 3.2 are about questions related to this topic.

6 Do people react to a stimulus more quickly with their writing hand than with their non-writing hand? (See Activities 3.4, question **3** and 4.11.) Alternatively, are reaction times of young people and older people different? (See Activity 3.10, question **1**.)

7 Investigate stereoscopic vision. (See Activity 3.4, question **4**.)

8 Investigate the use of non-verbal communication by different people. There are many suggestions about how to do this in Activity 4.18.

9 Investigate the probability of a drawing pin landing point upwards when it is dropped. How is this probability affected by the surface on which the pin lands? Activities 4.9 and 4.10 look at questions related to this investigation.

10 On which wrist do people prefer to wear their wristwatch? See Activity 4.17, question **1**.

11 Investigate when people are born. See Activities 5.1 and 5.4 for some ideas about this, but other questions can be investigated as well.

12 How many service tills are needed at different times of day in a post office or a supermarket? Some suggestions for studying problems like this are given in Chapter 6, and you will find some ideas about how to model shoppers arriving at random in Activities 5.2 and 5.5.

13 Are people able to persevere at a simple task for longer if they are alone, or with another person? (See Activity 3.10, question **3**.)

14 Are some groups of people quicker at spatial tasks than others? (See Activity 3.10, question **4**.)

How people use statistics

Here are examples of some statistical investigations that will be used to illustrate the discussion in later chapters.

Example 1 The national shoppers' survey

A questionnaire is delivered to all homes in selected areas, offering to recipients who return the questionnaire within 14 days 'a FREE Reward Pack of money-off coupons for leading grocery and household goods'. The aim is to gain market research information about which products people use every day, but the information given in the replies may be passed on to 'other reputable companies who want to send you offers that relate to your interests'. In other words, the returns also provide names and addresses to companies wishing to target their advertising on people with a particular interest in their product.

Example 2 The survey 'British Young People in the 80s'

This survey was commissioned in 1981 by a group reviewing the Youth Service in Britain. The objectives were to describe how young people spent their leisure time, and to investigate the reasons for their choices, in order to be able to plan the future provision of activities and resources for young people. Before the study could begin, the terms in the objectives needed to be defined more precisely. Young people were defined as aged between 14 and 19 years. Then the study focused on issues such as

- which leisure time activities or experiences young people were aware of, and which they tended to choose
- what relative value was assigned to these activities in terms of enjoyment, relevance to adult development, and other criteria
- how far these activities gave opportunities to exercise responsibility
- how important responsibility was to these people
- which activities were perceived by the young people as being provided or assisted by the Youth Service
- to what extent financial considerations encouraged or inhibited their choice of provision.

The survey was carried out using a questionnaire designed after an extensive series of interviews with young people. To ensure that their answers were not affected by any inclination to seek approval or disapproval of adults, the young people selected for the sample completed the questionnaire alone. The sample of 635 people was designed to be broadly representative of all young people in Britain in terms of age, sex and social class. But to obtain enough data from ethnic minorities and the unemployed, these groups were deliberately over-represented in the sample.

Example 3 Survey of pesticide usage on arable crops

The Ministry of Agriculture in England and Wales regularly surveys the use of pesticides on arable farm crops including wheat, barley, oats, rye, potatoes and peas. The information is used to help the government to advise on the use of dangerous chemicals. The samples are selected by using information from the agricultural census, so as to be representative of the area of ground covered by different arable crops in the six regions of England and Wales. A letter explaining the purpose of the survey is sent to the occupier of each farm selected for the sample, and each farmer later receives a personal visit from an experienced surveyor. The questionnaire is completed by the surveyor during an interview with the farmer.

Example 4 Pollution and hay fever

Researchers at St. Bartholomew's Hospital, London, noticed levels of hay fever in

cities similar to levels in the country. This seemed strange, since hay fever is supposed to be due to pollen in the air, and the level of pollen is much higher in rural areas than in towns. It was suggested that the people in city areas might be suffering from sneezes because of the higher levels of exhaust fumes. In order to test this hypothesis, researchers exposed patients to nitrogen oxides, ozone and other exhaust-related chemicals for up to four hours in sealed chambers. Afterwards they were exposed to pollen and other allergens. After they had been exposed to exhaust fumes, patients' levels of sneezing were, on average, three times higher than normal.

Activity 1.2 Why people collect data

In question 2 you are asked to find out how some published statistics are used, and you may have to wait for other people to give you information. Do not wait for this information before going on to Chapter 2.

1 **a** Pick one of the examples given above. As precisely as you can, define the question to be investigated.
b What data would you want to find or collect? How could you get this data?
c How would you analyse the data to help you answer the original question? What use would you make of graphical and numerical methods of analysis? What use of probability distributions?
d What sort of conclusions would you expect to be able to draw?

2 **a** Look through a recent copy of *Government Statistics: a brief guide to sources* published by the Government Statistical Service and available either from: The Library, Central Statistical Office, Cardiff Road, Newport, Gwent, NP9 1KG; or from Central Statistical Office Press, Publications and Publicity, Great George Street, London SW1P 3AQ. Choose one government publication which interests you. You may find a copy in a local library, or you could request a copy from the Central Statistical Office. Write a brief summary of the data it contains.
b Find out what you can about who would use these statistics, and for what purposes. You could contact the appropriate government department to ask them about their experience. You may be able to do this by letter or by telephone. Addresses and telephone numbers are published in many places, including *Government Statistics: a brief guide to sources.*

Reflecting on Chapter 1

What you should know

- stages in the statistical process
- some of the ways statistics are used.

Preparing for your next review

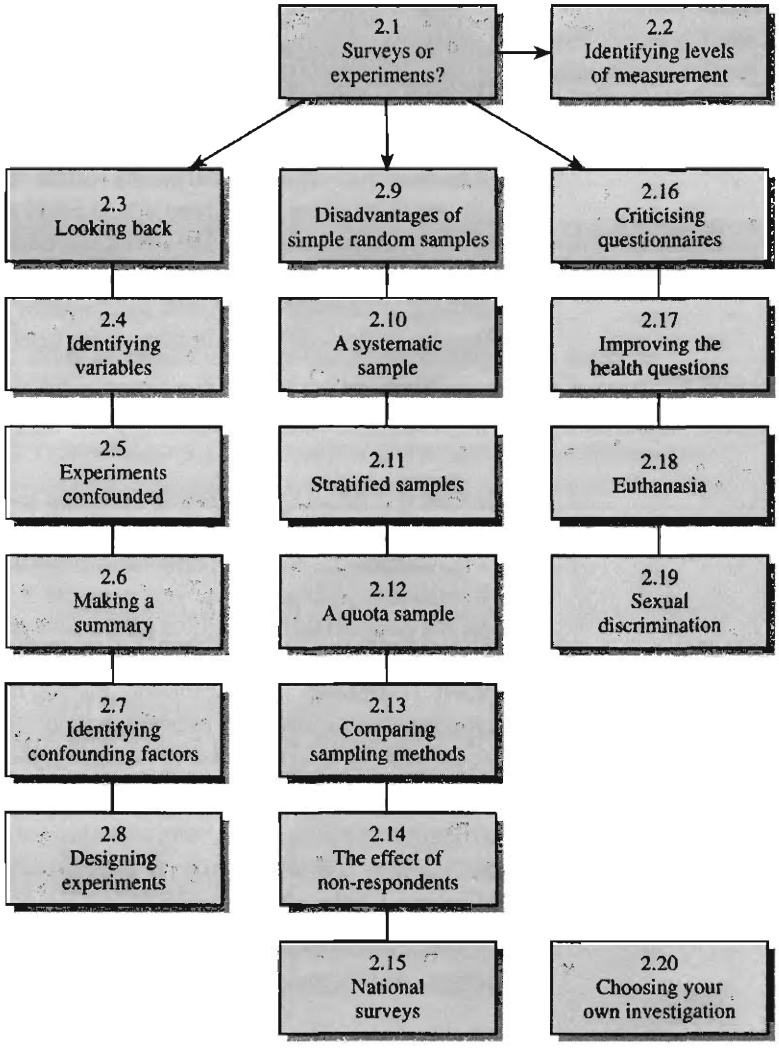
- Bring to the review your answers to Activity 1.2. Be ready to discuss them.
- There are no check questions for this chapter.

There are no practice exercises for this chapter.

2 Collecting real data

This chapter deals with some of the issues involved in planning a statistical investigation and deciding how to collect the data.

There are three main issues: first, the design of experiments, then the choice of sampling methods and finally the design of questionnaires.



Begin with Activity 2.1.

You can work Activity 2.2 at any stage.

Activities 2.3 to 2.8 deal with the design of experiments, and you should work them in sequence. If possible, work on Activities 2.5 and 2.6 with another student of statistics.

Activities 2.9 to 2.15 are concerned with sampling methods. Work them in sequence. If possible, discuss your answer to Activity 2.9 with another student of statistics. Activity 2.13 is a longer activity which requires a computer; you will find it helpful to share the work and discuss your results with other students.

Activities 2.16 and 2.17 are about questionnaire design; you should tackle them in sequence. They are not essential unless you choose to conduct an investigation which needs a questionnaire. You will need to discuss the questions in Activity 2.16 with other people, but not necessarily with other statistics students. Activities 2.18 and 2.19 give some ideas for special uses of questionnaires.

Activity 2.20 deals with your planning for your own investigation, and you should keep it in mind as you work through the chapter, although you will probably not be ready to tackle it until the end of the chapter.

Before you decide how to collect the data for a statistical investigation, you need to consider how you intend to analyse the results. You may not know precisely, but if you do not think about it in advance, you could end up having to collect more data than you need, which is time consuming. It is also important to identify the population in which you are interested. You have met the ideas of a population and a sample in your work on the units *Investigating data* and *Modelling with probability* in Book 2. The following definitions will remind you.

A **population** is the set of all the elements of interest in a particular study. It may consist of people, events, areas, animals, or whatever the items are that you are looking at.

A **sample** is a subset of a population selected to represent the whole population. Numbers, such as the sample mean $\bar{x}$, are called **sample statistics** and are often used to estimate the corresponding measures for the whole population. In order to have some idea of how reliable your estimate is, you will often need a distribution to model the population. You will therefore need to make assumptions about the structure of the population, such as whether the observations are independent, or normally distributed. The distribution that you use to model the population has parameters such as mean, μ, or average proportion, whose values are often assumed to be the same as the equivalent measures for the whole population.

Sometimes the whole population is not easily accessible, so you must limit your investigation to a subset of it. Although this restricts the generalisations that you can make, you may still wish to express your hypotheses and your results in terms of implications for the whole population.

Statistical investigations can be divided into two kinds according to the way in which you collect the data: either by experiment or by survey.

An **experiment** deliberately imposes a treatment on the subjects, or situation, to see the effect upon particular variables, which have been previously identified as being of interest. This may involve comparing a treated group with an untreated group to detect any differences. Observation of an existing situation does not count as an experiment. When an experiment is used, the researcher attempts to create the conditions to be investigated, so has much more control over the investigation. Experiments on human subjects are often difficult or impossible for both practical and ethical reasons. For example, it is difficult to imagine anyone conducting an experiment on humans to investigate whether smoking causes lung cancer.

In a **survey**, the investigator gathers information under existing conditions to find out about some characteristic of the population. For example, the early investigations into the effects of smoking on health were necessarily based on surveys. Data were collected about people who suffered from lung cancer, and it was found that many of them were smokers. However, this did not prove a causal relationship between smoking and lung cancer, since there might have been some other factor that both increased the likelihood of people becoming smokers, and also increased the risk of lung cancer. Other examples of surveys that involve sampling are opinion polls, and surveys to measure the audiences for the different TV channels. If a survey requires a sample, you need to think about how to select it.

In this chapter you will learn first about experiments, and later about sampling methods.

Activity 2.1 Surveys or experiments?

Consider these research questions.

1 How do people intend to vote in the next general election?
2 A biologist believes that the amount of fertiliser administered to plants may affect their growth, and wants to find out how much it affects growth.
3 Does anxiety improve people's performance in examinations?
4 Do people in the UK favour proportional representation?

For each of the research questions **1**, **2**, **3** and **4**, answer the questions **a**, **b**, **c** and **d** below.

a How could you collect the data to find an answer? Do you need to carry out a survey or an experiment?
b What is the sample in your chosen method?
c What is the population?
d In what ways might your results fail to be representative of the population?

Hints are provided for all of these in Chapter 8.

Measurement

Different kinds of data need to be measured or coded in different ways. These ways may restrict the methods you can use to analyse them.

Nominal scale of measurement

This is the simplest kind of data, in which people, animals or events are classified into groups. Although you use only simple statistical techniques on nominal data, remember that sometimes nominal data can be important explanatory variables. For example: gender, ethnic origin, religious or political affiliation, may be important in explaining people's behaviour. Numbers are sometimes used as labels for values of nominal data. For example, gender may be coded as 1 for male and 2 for female. The numerical order of the labels may not be significant, so you should take care not to use the values in calculations. It is usually meaningless to find the average gender.

Ordinal scale of measurement

When this level of measurement is used, items may be ranked in order of preference or order of occurrence. A familiar example is the grade given for effort on some school reports. One commonly used ordinal scale is the Registrar General's Classification of Occupations used to indicate social class

I	higher professional
II	professional
IIIa	clerical, white collar routine jobs
IIIb	skilled manual
IV	semi-skilled manual
V	unskilled manual.

There are problems with this method of scaling occupations, particularly in the classes IIIa and IIIb, and you may disagree with the hierarchy which it implies. In spite of these problems, it has yet to be superseded.

Notice that the differences between one class and the next are not equivalent, so it does not make sense to treat the values as numbers. You cannot speak of a mean social class, although it might make sense to talk about a median class.

Interval and ratio scales

An **interval scale** has equal intervals all the way along it: so the difference between 25 °C and 30 °C represents the same increase in temperature as the difference between 0 °C and 5 °C.

Ratio scales are interval scales that have an absolute zero. With ratio scales you can make statements using ratios such as

'he is four times as old as his son'
'she is twice the height of her child'
'this bag weighs half as much as that bag'.

Most interval scales are also ratio scales. An example of an interval scale which is not a ratio scale is temperature: a temperature of 20 °C is not twice as hot as 10 °C because zero is an 'arbitrary' value on the scale, not an absolute zero. Notice that the ordinal scale for social class is neither an interval scale nor a ratio scale. In general, the most sophisticated statistical techniques are used for data measured on an

interval or ratio scale. Nearly all of the interval scales used in this course are ratio scales. The hierarchy of scales is shown in Figure 2.1.

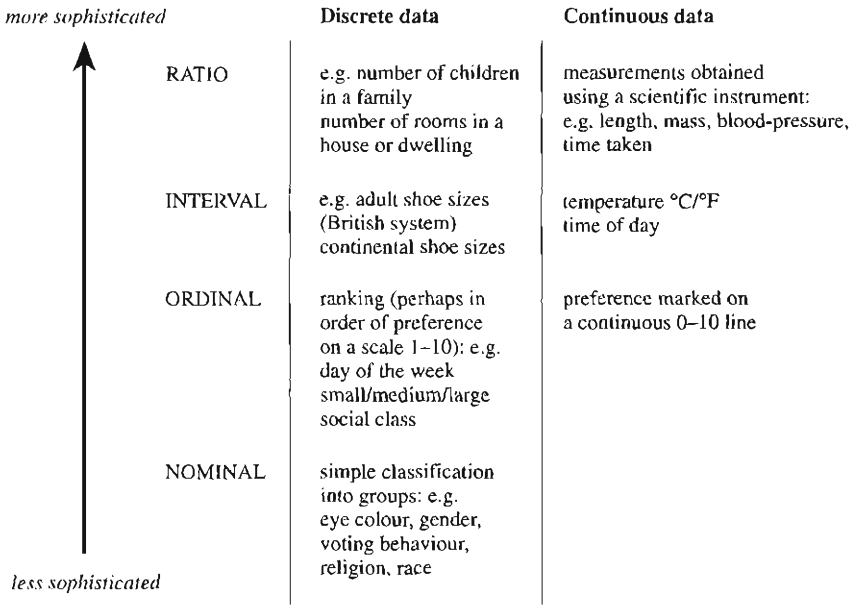


Figure 2.1

Levels of measurement

Figure 2.2 is a summary of the techniques which can be used on data from the different scales of measurement.

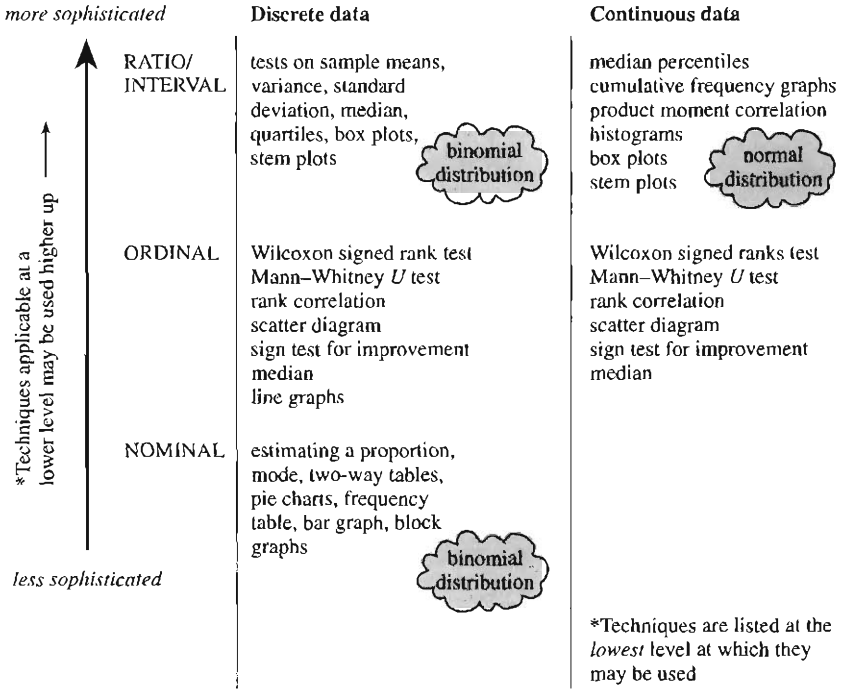


Figure 2.2

Statistical techniques and models

Activity 2.2 Identifying levels of measurement

Look at either the TOWNS dataset in Appendix A in the *Investigating data* unit in Book 2 or the HEIGHT dataset in the appendix to this book. Read the description of each variable used. Decide for each variable whether it is measured at the level of nominal, ordinal or interval/ratio data.

Experiments

An experimenter has to take into account all the factors that may vary in the situation being studied. These factors are called **variables**. It is important to distinguish between different kinds of variable.

Here are two examples which illustrate the definitions that follow.

Example 1

A biologist believes that the amount of fertiliser administered may affect the growth of plants, and designs an experiment in which the amount of fertiliser is strictly controlled and the resulting growth is monitored.

Example 2

A psychologist believes that people will find it easier to remember words than random strings of letters, so to test this hypothesis designs a memory experiment using strings of six letters.

Explanatory variables are also called **independent variables.**

Explanatory variables

Explanatory variables are the factors that the experimenter hopes to vary systematically so that the effects of changes in one may be observed in other variables.

In Example 1 the explanatory variable is the amount of fertiliser administered.

In Example 2 the explanatory variable is the presence of meaning in the strings of six letters that are presented to the people in the sample. If the string of six letters is not a word then it is nonsense.

Response variables are also called **dependent variables.**

Response variables

Response variables are the results of the experiment and must be measured carefully to assess the importance of any changes. When you vary the explanatory variables, you observe the changes in the response variables.

In Example 1 the response variable might be the height or the leaf size of the plants after a defined period of time. In Example 2 the response variable is the number of letters correctly remembered.

Controlled variables

The experimenter is concerned to investigate the effect of the explanatory variables, so any other variables that might cause changes in the response variables need to be **controlled**. Ways of controlling variables are described below.

One way of controlling variables is to exclude them. For example, you can remove gender as a source of variation by including only females in a sample. Sometimes you can hold other variables constant; for example, you can keep temperature, background noise, surroundings, instructions, and so on, identical for all subjects. But controls such as these are likely to restrict your ability to make generalisations, so you need to be careful when you interpret your results.

Sometimes you can control variables by randomisation. This is explained in more detail later in the chapter. If you think that the order in which subjects undertake particular tasks in the experiment may be important, you can randomise the order in which they perform them.

In some situations, the only way to control some variables is to measure them. For example, in an agricultural experiment, you can only control the temperature and rainfall in this way.

Extraneous variables

Extraneous variables are uncontrolled variables that may influence the outcome of the experiment. They may include the subjects' level of motivation, state of health, and so on. Sometimes the experimenter may not even be aware of them.

Figure 2.3 shows how the different kinds of variables are related to one another.

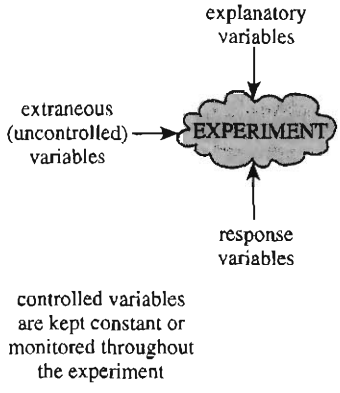


Figure 2.3
Variables in an experiment.

Activity 2.3 Looking back

You may find the diagrams provided in the hints section (Chapter 8) helpful for organising your answer to this activity.

H **1** Look back over the explanation of experimental variables, and write a paragraph about each of Examples 1 and 2 to summarise the variables in each experiment.

Activity 2.4 Identifying variables

1 For each of the two experiments described below, identify, if possible,

- the explanatory variable
- the response variable
- the controlled variables
- the extraneous (uncontrolled) variables.

a A student decides to investigate whether fire retardant sprays are effective. He collects samples of ten different fabrics with different fibre content (for example: rayon, wool, polyester, linen, nylon, cotton and mixtures of these). He cuts two 10 cm squares from each piece of fabric. One square of each type is sprayed with the fire retardant spray and allowed to dry. He uses a Bunsen burner and a stop-watch to carry out an experiment in a fume cupboard to see how long each piece of fabric takes to burn completely.

b An experiment is set up to see whether people can improve their IQ scores with practice. A group of students is given a similar IQ test each week for six weeks. The scores on the first and last tests are compared.

2 Explain why the following two studies are not true experiments.

a To investigate whether the consumption of alcohol by drivers increases road accident rates, information is collected from official sources, and the proportion of accidents involving drunken drivers is calculated. Comparisons are also made of accident rates before and after the introduction of stricter legislation on drink and driving.

b A dental researcher collects data from dental records. She is interested to see whether the application to teeth of a protective coating (called fissure sealant) helps to protect them from decay. She examines the dental records of eight-year-old children registered at two similar dental practices. She compares the number of fillings in children's teeth at the first practice, where fissure sealant is routinely used, with those for the other practice where fissure sealant is not used.

3 You may have conducted a 'taste test' when you studied the binomial distribution. Describe how you would set up a taste test to see whether people can distinguish between bottled water (not carbonated) and tap water. Which variables would you need to control?

'Confounding' is a technical word in statistics, and is not necessarily a bad thing in an experiment.

Things that can go wrong in experiments

The effects of explanatory or extraneous variables are said to be **confounded** when it is impossible to isolate the influence of either one of them on the other. This section deals with issues related to confounding.

Activity 2.5 Experiments confounded

If possible, discuss these experiments with another statistics student.

1 For each of the following experiments, consider what might go wrong and

whether you could deal with it.

a A researcher wishes to see whether the introduction of a weekly exercise programme will raise the self-esteem of a group of elderly residents in a home.

b A study is to be carried out to investigate whether regular doses of vitamin pills for children can affect their progress at school.

c In an experiment to test which of two types of running shoe is better for 100 m sprints, each person in a group of athletes is asked to run 100 m, wearing first one type of shoe and then the other.

As you read the following text, keep in mind the experiments described in Activity 2.5.

The effect of the experiment itself

The very fact that a group of people has been selected for an experiment can make them feel 'special'. Students selected for a new method of teaching may work harder (in the beginning at least!) while their teacher, believing that the new method is better than the old, and pleased to have been asked to take part in the experiment, may well work harder too.

In the old people's home in Activity 2.5, factors other than exercise could improve the self-esteem of the residents. The effect of the exercise programme may be confounded with the effect of the experiment itself. Participants will be interviewed individually at the start and end of the programme, and will also have increased social contact with each other during the exercise sessions. They may enjoy these social aspects so much that their self-esteem improves, irrespective of any improvement in suppleness and physical capabilities.

In this situation, what is needed is a **control group** which is as similar as possible to the experimental group. The control group could receive two individual interviews over the same interval of time. They should also meet weekly as a 'coffee and chat' group so that they can be compared with the experimental group. In the case of the school class taking vitamin pills, it is likely to be much more difficult to design an experiment which includes a control group.

Be wary of any studies conducted using volunteers. Volunteers are not a random sample. They may have special qualities which make them different from non-volunteers. Are they more adventurous, more extrovert, more motivated, more foolhardy?

Placebo effect

A similar effect is the placebo effect, which occurs in clinical trials of drugs and other medical treatments. Many patients respond positively to any treatment, even a dummy medication such as a sugar pill. This recovery is not just a mental phenomenon; physical improvements can actually take place.

A study to investigate the effect of children taking vitamin pills on their progress at school would need a control group taking 'dummy' pills and, ideally, a second control group taking no pills at all.

Experimenter effect

In experiment **b** on the previous page, the teacher, knowing which children have received the genuine vitamin pills, will expect improvements in those children's performance at school. People quickly respond to positive expectations and encouragement and will perform better, even if this is not expressed outright. This is an example of the experimenter effect.

Many studies of this type are set up as **double blind** experiments so that the doctor, or person administering the experiment, does not know which patients are receiving the dummy pills and which are the real pills. This is to overcome the experimenter effect.

Maturation

If an experiment takes place over a period of time, the subjects may well change anyway, even without the effect of the experiment. For example, some children's rate of progress at school may improve as they mature, irrespective of whether they receive vitamins.

So all experiments should include a control group to reflect changes which are taking place naturally, or due to other outside influences.

Order effects and practice effects

In the experiment to test two types of running shoe, you may suspect that the athletes' second run would be slower if they are tired from the first run. On the other hand, if you think they have just 'warmed up' on the first run, you might expect the second to be even faster. To overcome this, you could insist on a rest period between the two runs.

If you are still worried about an order effect of this kind, you can randomise the order in which the shoes are worn. One way is to divide the sample of athletes into two groups at random, and ask the first group to run first in shoe A and then in shoe B, but ask the second group to run first in shoe B and then in shoe A. In this way, order effects should average out over the whole sample.

In experiments where subjects have to repeat similar tasks, such as the memory test outlined on page 34, their performance could improve later in the experiment, as they gain more practice. For this reason, you need to randomise the order of the words and strings of letters so that any practice effect is averaged out over a number of subjects.

You can use randomisation to average out such effects in an experiment. You should also use it to allocate subjects to groups – that is, to the experimental group and control group – or to different treatments within the experiment. This technique will average out extraneous factors which have not been controlled in the design of the experiment.

Activity 2.6 Making a summary

If possible, work with another student for question 3.

1 Make brief notes about the meaning of the following terms:
a experiment effect **b** placebo effect **c** experimenter effect
d maturation effect **e** order effect.

2 Describe briefly why you would use
a a control group **b** double blind trials **c** randomisation.

3 Look back at Activity 2.5, and decide how you could now improve your answers. If possible, discuss your ideas with another student.

Activity 2.7 Identifying confounding factors

1 For each of the following experiments, **a** to **d**, answer the question asked, and decide whether there are any confounding factors which you would need to consider in presenting the findings of the study. Identify the explanatory and response variables and say which extraneous variables are confounded with the explanatory variables.

a A researcher in education asks for local teachers willing to try with their pupils her new method of teaching reading. She finds ten teachers of year 1 willing to do this. For comparison, each of these teachers is paired with another teacher using a traditional method of teaching reading at the same, or a similar, school. At the end of one year, she compares the reading scores of the children taught by both methods, and discovers that the children who used the new method have an average reading score 10% higher than those who used the traditional method. How could you decide whether her new method is better than the traditional method?

b A school nurse keeps records of the illnesses of the children in her care. One year she finds that the year 7 pupils missed 10% of their class time through illness in the winter term. The next year she arranges for the same group (now in year 8) to take a multi-vitamin tablet each day. She notes that their absence rate has gone down to 5%. What further information do you require to decide whether this is conclusive evidence that multi-vitamin tablets are beneficial?

c A psychologist wishes to see whether people who are good at maths are better at memory tasks involving numbers rather than letters. He selects a random sample from students taking science A-levels. The students first take a test which requires them to remember random groups of letters, and then a second test in which they are asked to remember random strings of numbers. On average they score higher on the number memory test. Does the experiment provide conclusive evidence?

d A science teacher who has made a series of videos with the aim of raising his students' awareness of ecological issues, wants to find if the videos are effective. He administers a pencil and paper test to a group of students. They then watch the videos over a period of six weeks. At the end of the course, they take a similar test which shows that their knowledge has improved and their attitudes towards conservation issues are much more positive. What more do you need to know before you can judge from this evidence that the videos are effective?

2 A survey of doctors finds that some of them would give a placebo to a patient complaining of a pain or symptoms for which the doctor can find no cause. If the patient's symptoms vanish, these doctors conclude that there was no real physical

cause for them, since no real treatment has been given. Is this conclusion valid? Is this a correct interpretation of the placebo effect? Justify your answer.

Types of experiments

Independent samples

Samples are **independent** if they are selected on the basis of a strictly random procedure. This is the simplest design for experiments; it can be used in laboratory experiments and with human subjects in or out of the laboratory.

It is hoped that randomisation will ensure that, by chance, human subjects are fairly well matched across the samples, in terms of gender, levels of intelligence, attentiveness, creativity, and so on. Randomisation cannot ensure that the two samples will be perfectly matched on any one extraneous variable, but there will not be any systematic bias in favour of any one subgroup. In the long run, if the experiment were repeated indefinitely, with random allocation each time, all the differences would balance out.

If you are using human or animal subjects, this design works best if you can use large samples. If you are only able to use a small number of subjects, any imbalance between groups may considerably influence the results and cause the explanatory variables to be confounded with extraneous variables.

Matched groups

You may be able to use matched groups with human subjects in an attempt to eliminate as far as possible the effect of extraneous variables. The best studies of this kind have used identical twins, but even in these cases the subjects are matched only for genetic and physical characteristics. Personality traits may be quite different! It is extremely difficult to select pairs of individuals who are very similar, even if you restrict yourself to considering a limited number of variables that you consider to be important. Even if you can identify the most influential variables, it is a difficult and time-consuming task to recruit pairs of subjects who are closely matched. Consequently, people seldom use this design except in twin studies. It is likely to be very difficult for you to use this design on human subjects in your own investigation.

On the other hand this design might be used in a chemical study. In an investigation of crystallisation, for example, in order to match the conditions of temperature, air pressure, humidity, vibrations, and so on, the researcher might prepare several samples of solution simultaneously, each containing a different concentration.

'Before and after' design

'Before and after' design is also called **repeated measures design.**

Very often, the only way in which you can be sure that two groups of human subjects will have identical characteristics is to use the same subjects in each group. You must test or measure the subjects before and after the treatment – which might be, for example, a course of lessons or a course of medical treatment. You can

compare the two measurements and estimate the effect of the treatment. This design is particularly useful when the changes brought about by the experiment are likely to be very small. You can isolate these small differences in the statistical analysis, eliminating or ignoring the differences between subjects. When you use human subjects in an experiment, this design of experiment is often the most efficient that you can use, because it minimises the variability in the results. However, maturation, and order and practice effects, may still present difficulties.

Activity 2.8 Designing experiments

Use a simulator instead of letting your subjects loose on the road!

Here are three statements.

1 Driving skill varies as a function of the level of alcohol in the blood.

2 People can identify more rapidly that two letters have the same name if they are written identically than if one is a capital and the other is a small letter.

3 People weigh more last thing at night than they do first thing in the morning.

Outline a simple experiment to test each statement. State:

You can refer to the *Distributions* unit in Book 5 if you need a reminder about the null and alternative hypotheses.

a your null hypothesis

b your alternative hypothesis

c your explanatory variable

d your response variable

e whether you would use independent samples, matched samples, or the same subjects in both conditions

f which variables you would control

g which extraneous variables would be controlled by randomisation.

Random sampling

In the *Investigating data* unit in Book 2, you learned that the method which you use to select a sample is critical. In most situations it is completely impractical to survey the entire population, so it is very important to be able to obtain a sample which has the same characteristics as the whole population. Such a sample is called a **representative sample**.

Simple random sampling methods ensure that every member of the population has an equal probability of being chosen, and no subgroup within the population is deliberately excluded. This ensures that any deliberate source of bias is eliminated, so the sample is likely to be representative.

However, as you saw in your work with probability simulations in the *Modelling with probability* unit in Book 2, any one random sample may, by chance, appear to be unrepresentative. The unbiased nature of random samples emerges only in the long run, when a large number of random samples is observed. Nevertheless, the

Methods of selecting a simple random sample from a whole population, were described in the *Modelling with probability* unit in Book 2.

likely extent of any random fluctuations can be predicted and modelled mathematically. For example, you can calculate the variability in estimates from random samples, as you found in the section on estimating proportions in the *Distributions* unit in Book 5. These calculations cannot easily be made for non-random samples.

When population members present themselves one at a time as potential members of the sample, you can obtain a **simple random sample**, even when you have no list of the population. You can adopt a randomisation procedure using random numbers.

Activity 2.9 is designed to help you to prepare for the next section.

Activity 2.9 Disadvantages of simple random samples

If possible, you should discuss your answer to question 2 with another statistics student.

Think again about these two problems.

- How do people intend to vote in the next general election?
- Do people in the United Kingdom favour proportional representation?

1 Why might a simple random sample be inappropriate for an opinion poll investigating these questions?

2 Write a paragraph to describe how an opinion poll organisation could obtain representative samples to answer these questions, without going to the time and expense of finding and interviewing a sample that really is a simple random sample.

Non-random sampling

Instead of simple random sampling, various other methods of sampling are sometimes used. They may have practical advantages over simple random sampling because they are quicker and easier to carry out. On the other hand, all have at least one disadvantage. You cannot easily predict the distribution of any sample statistics from a non-random sample, or model them mathematically. For example, it is not known whether the distribution of sample means based on a non-random sampling method will form a normal distribution. This may make it difficult for you to interpret the results of an investigation, and you may not be able to give estimates in the form of confidence intervals.

Systematic sample

You could draw a **systematic sample** by selecting every tenth (or twentieth) item from the population, starting from a randomly selected point in the first ten (or twenty).

If the population members are in random order, then a systematic sample is equivalent to a simple random sample. An alphabetically arranged list is not in random order, but this may not matter if you have good reason to believe that the variable under consideration is distributed randomly throughout the list.

Systematic sampling is convenient and cheap. For instance, it is very easy to sample every tenth person entering a library, leisure centre or hospital casualty department. It is an especially convenient way to sample from a list.

One difficulty with systematic sampling is that you may have no way of knowing whether there is any cyclical variation in the sequence or list that could cause bias. For example, if you select every tenth house number in a street, then all the houses chosen may be on the same side of the street, which might affect the results of the survey.

How could you avoid such a bias?

However, the advantages of this method are so great that such anxieties are often unimportant.

Activity 2.10 A systematic sample

The Department for Education monitors in some detail the progress of a sample of children through the education system. Each school is asked to provide information about the public examination results of all students in each year group who were born on either the 5th or the 25th day of any month. This has the benefit that the same students can be followed through from year to year.

1 Is this a systematic sample? Justify your answer.

2 Is the sample likely to be representative of each year group?

3 What proportion of the population is in the sample? What assumptions have you made in your calculations?

Stratified sample

It is possible that a particular simple random sample might not be representative of the population. For example, a random sample of 30 students selected from a mixed sixth form college could consist entirely of male students. It would be preferable for the gender ratio in the sample to be similar to that in the college as a whole. Other important factors might be: courses being followed, social class or ethnic origin.

If you have information on variables that may be relevant to your investigation, such as gender, social class and so on, you can use this to divide the population into subgroups or strata. Then you can select a simple random sample from each subgroup. This ensures that the sample represents the population with respect to these variables.

If you use this scheme, you need to decide how many sample units to select from each subgroup. One approach is to adopt proportional representation, so that if 55% of students are female, you ensure that females also form 55% of the sample.

On the other hand, if some subgroups are particularly interesting to your research, or if you think that their replies are likely to be more variable than those of other groups, you might wish to have them over-represented in the sample, even though they may form a small proportion of the population. This means that, to estimate a mean or proportion for the population, you will have to re-weight the survey results after analysing the sample.

The survey British Young People in the Eighties is described at the end of Chapter 1. The survey was stratified using age, sex, social class, ethnic origin and employment status. It was recognised that each of these factors could be associated with differences in attitude and leisure behaviour. It was considered particularly important to obtain good data about ethnic minorities, so the proportions of people of Asian and West Indian origin in the sample were deliberately increased. Similarly, the proportion of unemployed people aged 17 to 19 was also increased.

Activity 2.11 Stratified samples

1 A council planning officer needs to estimate how many children under 16 live in three high-rise blocks of flats. Each block of flats has 18 floors and there are 14 flats on each floor except the top floor, where there are only 6 luxury penthouse flats. The officer has no precise information on the size of family at each address, but believes that most of the larger families will be living on the lower floors. The number of children in each family may be extremely variable, ranging from zero to as high as ten.

H How could the officer's beliefs be used to construct a stratified sample of 50 families from the flats? Consider what should be done when an enquiry produces no response. Why would a stratified sample be preferable to a systematic sample or a simple random sample in this case?

Quota sample



This method of sampling is often used in street interviews for market research surveys and opinion polls. It is easier and cheaper than sending interviewers to randomly selected homes. A quota sample is, in effect, a stratified non-random sample. First, you need to specify the strata and decide how many people to select from each. This gives you the quotas for interviews.

To choose people to fill their quotas, the interviewers must go to a suitable place, such as a shopping centre on a busy street. For example, the quota might include three women aged 18–24, four men aged 25–44, and so on. The interviewer must continue looking for women aged 18–24 until she has interviewed exactly three of them, but she must stop interviewing men aged 25–44 as soon as she has found and interviewed exactly four.

There are some problems for interviewers who must subjectively estimate the age, and sometimes social class, of people approaching them. If an interviewer begins an interview and subsequently finds that the person belongs to a group whose quota is already met, the interview must continue to a suitable stopping point so as not to cause offence, but the result is discarded later. It may be quite difficult to find people in certain categories. For example, if the survey takes place during normal working hours, it may be hard to find the required number of office workers. Since interviewers have to judge subjectively the age of potential respondents, the final sample tends to cluster in the middle of each age range. There is also some evidence that interviewers are more likely to approach people who look friendly.

Quota sampling is very simple to use, as no list of the population is needed. On the other hand, because no list is involved, it can sometimes be difficult to define clearly the target population: that is, the population from which the sample is really drawn. Quota sampling is also very much cheaper than simple random sampling, since it avoids the need to call back to find selected sample members who are unavailable, and it ensures that those included are willing to participate. Travelling for the interviewers is minimised, and results can be made available very quickly. But those selected will tend to be the most willing to participate, and this may introduce some unintended bias into the sample.

Activity 2.12 A quota sample

The survey British Young People in the Eighties, described at the end of Chapter 1, used a method of quota sampling.

1 Consider the difficulties involved in obtaining a sample by any other means, and write a brief summary of the factors which led the researchers to use quota sampling to obtain their sample.

2 What possible bias might be introduced into the sample because it is a quota sample? Write a brief account of the factors for which you would look out when interpreting the results.

Cluster sample

Suppose that you want a sample of school pupils representative of pupils in state schools in the United Kingdom. You could select a random sample of local authorities, and then a random sample of schools within those authorities, and even a random sample of classes within these schools. This has the advantage that interviewers do not have to travel over large areas, since members of the sample are concentrated in a relatively small area. On the other hand, there may be problems if the initial choice of local authorities is not representative of the country as a whole.

This procedure is called **cluster sampling**, and can be used in situations where there is a hierarchy of units, such as local authorities, schools, classes and pupils.

Activity 2.13 Comparing sampling methods

Use a statistics package or a spreadsheet to do this activity. Your teacher has some notes about some statistics packages and spreadsheets.

If possible, you should share the work in this activity with another statistics student. You will then be able to discuss your results from questions 7, 9 and 10.

1 Turn to the HEIGHT dataset in Appendix A (Chapter 9). The value –1 has been used as a code for missing values in this dataset.

Select ten simple random samples of 20 husbands from the HEIGHT dataset, and calculate the mean height of husbands in each sample.

2 Split the individuals on the HEIGHT dataset into two groups according to the age of the husband:
a age 34 and under (that is to say, born since World War II, since the information was collected in 1985)
b age 35 and over.

How many men are there in each category? What proportion of the total appears in each age group?

3 Use the strata defined in question **2** to select ten stratified random samples of the heights of 20 husbands. Find the mean height of each sample.

4 Now use the strata and proportions in question **2** to select ten quota samples of 20 men from the HEIGHT dataset. Find the mean height of each sample.

5 Select ten systematic samples of 20 men from the HEIGHT dataset, and find the mean height of each sample.

6 Choose a cluster from either column 7 (tenancy) or column 2 (social class). Find the mean height of a random sample of 20 men selected from that cluster. Repeat this until you have done ten cluster samples.

The difficulty in this case is that there are several missing values in column 2 (and in column 7); these rows will be omitted from all your samples.

7 Collect together your results for the sample means for:
a simple random samples
b stratified random samples
c quota samples
d systematic samples
e cluster samples.

8 Draw up frequency tables and histograms for your sample means.

9 How do the results from different types of samples compare?

10 Compare your results with the actual mean height of husbands in the whole HEIGHT dataset.

Response rates

In any survey of people, however well you design the sample, there are always some people who either cannot be contacted, or who refuse to take part in the survey when you do contact them. These non-respondents usually differ from people who respond to surveys; for example, in a survey about people's attitudes to women managers, those in senior management might be too busy to respond, but their attitudes might be very different from those who do respond. It is therefore important, when interpreting the results of a survey, to note what proportion of the original sample were interviewed and included in the results. This is the **response rate**.

It is sometimes not obvious how you should calculate a response rate. Some items in the survey questionnaire may have a lower response rate than others, so a strict

definition of the response rate for the whole survey would include as respondents only those from whom every piece of information was collected.

In the survey British Young People in the Eighties, there was a difficulty with response rates. The response rate for an ethnic group was calculated as the percentage of participants who completed their questionnaire, but this definition did not show the full picture. The researchers had some difficulty recruiting Asian participants, since the parents of these young people often tended either to veto their children's participation altogether, or wished to reserve the right to check the responses that their children might make. Some even offered to complete the questionnaires themselves on behalf of their children. Since the purpose of the study was to seek the views of young people themselves, neither of these requests was acceptable. There may, of course, have been other parents who completed the questionnaires on behalf of their children, without telling the researchers what they had done. This possibility would be a source of error in the responses, but provided that it did not happen often in any one subgroup of the sample, it need not be critical to the results.

Activity 2.14 The effect of non-respondents

If possible, discuss these questions with another statistics student.

1 In the General Household Survey conducted by the Office of Population Censuses and Surveys, an analysis of the census returns of non-respondents revealed that older people and households with no dependent children had much lower response rates than the average. Why do you think this was? Do you think this finding should influence your interpretation of the results of the survey?

2 Look back to the survey problems described at the end of Chapter 1 (page 4). Choose two of the surveys, and try to predict what kind of people would be non-respondents. Do you think these non-responses could significantly influence the results?

Activity 2.15 National surveys

If possible, discuss these questions with another statistics student.

Investigate one of the following. If possible, find out who carries out the research, what sample they use, how the data is collected and what the response rate is. Discuss with another student the possible sources of error.

- General Household Survey (Office of Population Censuses and Surveys – OPCS)
- New Earnings Survey (Department of Employment)
- National Child Development Study (Department of Social Statistics, City University, Northampton Square, London, EC1V 0HB)
- Family Expenditure Survey (OPCS)
- Survey of Personal Incomes (Inland Revenue)
- British Social Attitudes Survey (Social and Community Planning Research, 35 Northampton Square, London EC1B 0AX)
- Census of Population (OPCS)

- OPCS Longitudinal Study
- England and Wales Youth Cohort Study (Department of Employment).

Questionnaire design

All questionnaires need careful planning. Therefore, you should always aim to try out a first draft of your questionnaire in a pilot survey before you conduct your full survey. The pilot is usually carried out on a few people, and can be used to find out what ways of wording your questions work best. It also gives you the opportunity to find out how long it will take respondents to complete the questionnaire, which may lead you to reduce the number of questions.

Before you write any questions, think carefully about what you want to find out, and write down your hypotheses. Decide what information you need, and ask only those questions that are necessary. Eliminate all the rest. If you ask people to answer too many questions, they may give up, or fail to answer conscientiously. Remember that people are giving up their time to help you.

A questionnaire may be administered in an interview, where each person is asked the questions by an interviewer, but it is usually more convenient if it can be handed out on paper for people to make written responses. If you use an interview, you may need to investigate whether the age and gender of the interviewer are likely to affect people's responses.

Activity 2.16 Criticising questionnaires

If possible, discuss the questions in this activity with other students. It doesn't matter whether or not the other student is studying statistics or even mathematics.

1 Discuss the following with other people.

a Have you ever been given a questionnaire to fill in which you found difficult to complete, or thought was not fair, or did not enjoy filling in?

b What was it about the questions that made you feel like that?

c Write down as many points as you can think of that need to be considered in writing good questions for a questionnaire.

2 Here is part of a questionnaire which was designed to survey the awareness of health issues among people aged 16 to 19. Each question is unsatisfactory in some way. Explain what is wrong with each question.

a How much exercise do you get?

- above average
- average
- below average

b Have you brushed your teeth today?

c Are you an unhealthy person?

d Do you ever have headaches?

e Have you suffered from backache or a stiff neck recently?

f As passive smoking can be very dangerous to health, do you agree that smoking should be banned in public places?

g Would you prefer not to use a non-medicated shampoo?

h How far do you walk each day?

i How much alcohol did you consume in the last month?
j What percentage of your daily diet consists of
- carbohydrates
- fats
- proteins?

k Do you think that individual people are alienated by the bureaucratic structure and procedures of the modern Health Service?

You can also use the pilot survey to see if different ways of wording questions affect people's understanding, or the answers they give. You may need to investigate whether the age and gender of the interviewer affect people's responses, or whether people tend to pick the first and last items on a list of multiple responses. If you ask for a numerical estimate, do people tend to choose a number close to the average or towards the extremes?

Factors to consider when designing a questionnaire

1 Open questions

Open questions allow the respondents complete freedom to answer as they please. Unlike closed questions, they do not offer a list of alternative answers or boxes to tick. Instead the respondents can express their own thoughts and ideas in their own language. Answers may be spontaneous and perhaps not well considered.

Open questions are easy to ask, but some people find them difficult to answer, and they can be difficult to analyse. If you wish to do any statistical analysis, you will need to code or classify the answers in some way. For this reason, this type of question is useful in a pilot survey to help establish the range of answers to offer in a closed question.

In your pilot survey, open questions may also help you to generate hypotheses.

2 Closed questions

Closed questions are quicker to answer as the respondent only has to choose from a list provided. They can also be analysed more easily, as the responses are already coded. However, the answers may lack spontaneity, and the list may not allow respondents to answer as they wish. Your list of alternatives needs to include every answer a respondent may wish to give, so it is often desirable to include an open category of answer labelled 'other'.

These questions need to be carefully tested in your pilot questionnaire.

3 Leading questions

Leading questions suggest a particular answer and may indicate the questioner's point of view. Examples are: 'Are you against giving too much power to the trade unions?' and 'Most people think that ... do you agree?' Even a question that appears neutral like 'Do you read a daily newspaper?' suggests that this is something that people should do; 'Do you ever read a newspaper?' may be better.

Wherever possible, try to use a neutral question, or one that contains two opposite views for the respondent to choose from.

4 Ask one question at a time

Do not ask questions like 'Do you suffer from headaches or sickness?' If the respondent says 'Yes' you do not know to which symptom they are referring.

Try to avoid general questions such as 'What do you think about the college canteen?' You will usually obtain more useful responses from a series of specific questions, such as: 'What do you think about the choice of main course meals served by the college canteen?'

The question 'Do you often visit the cinema?' is too vague, and is difficult to answer. 'Have you been to the cinema in the last four weeks?' will give people a specific time scale.

5 Negatives

Negatives often cause confusion. The question: 'Do you agree that smoking should not be allowed in public places?' may be better rephrased. To answer this question, the respondent has to say 'Yes' if they think smoking is a bad thing, and 'No' if they think it is all right.

Double negatives should always be avoided. This question shows the confusion that is generated: 'Do you not think that people should not be allowed to consume alcohol in the street?'

6 Ambiguity

Some words have different meanings to different people according to their gender or social class. For example, consider the meals called 'tea' and 'dinner'; think about the times when these meals might be eaten, and the type of meal that might be eaten. Many school children refer to their mid-day meal as 'dinner'.

In a housing survey it was found that respondents reported fewer bedrooms than their houses actually contained. This was because they did not consider a study, play room, sewing room or even a guest room as a bedroom.

When you read through your questionnaire, you know exactly what you intend the questions to mean. It is difficult to anticipate whether other people will understand the questions in the way that you intend. Even before your pilot survey, it is worth asking a few people to read each question, and tell you what they think you are asking.

7 Respondents' knowledge and competence

Only ask questions which people are likely to be able to answer, particularly if you are concerned with facts rather than opinions. Avoid asking questions outside your respondent's experience. Few 18-year-olds are likely to have an opinion to express on 'Do you think the government should make greater efforts to reduce mortgage interest rates?'.

Do not rely on what people remember over a long period. People might just be able to tell you what they ate and drank yesterday, but would be unlikely to remember accurately what they ate and drank on previous days. If you are interested in routine activities such as eating, smoking or watching TV, you need either to limit the time scale to one day, or to ask people to keep a diary for you.

Other behaviour such as using sports facilities, going to the cinema or theatre, may be remembered for a longer period of perhaps a month, but will be inaccurate for more than that.

8 Order of the questions

When planning the order of your questions, ask the easier, more factual questions at the beginning of the questionnaire. Leave more difficult or potentially embarrassing questions until later.

Sometimes you may need to ask respondents to admit to socially unacceptable behaviour. Leave these questions until later in the questionnaire, by which time you would hope to have established some rapport. Do not express any opinions or be critical in your questions. For some questions, respondents may need a guarantee of anonymity.

Activity 2.17 Improving the health questions

1 Look back at question **2** of Activity 2.16 and, for each part, identify which of the factors 1 to 8 described above apply. Try to write improved versions.

2 Identify two important health issues that have been missed out. Write questions to address them.

3 Try out your new set of questions on about ten other students.

4 In the light of the responses you get, consider what further improvements you might make to the questionnaire.

Summary of main points

- Plan your questionnaire.
- Keep questions to a minimum.
- Use simple language.
- Avoid leading questions.
- Do not use double negatives.
- Be unambiguous.
- Do not ask general questions if you want specific answers.
- Do not rely on people's memories over a long period of time.
- Try to avoid questions outside your respondents' direct experience.
- Be careful with embarrassing questions.
- Pre-coded questions make analysis easier but may force respondents to give the wrong answer.
- Pre-test your questionnaire.

Split polls

In situations where researchers suspect that either the wording of the question, or the order of the suggested answers, will affect their results, they may use a technique called a **split poll**.

This means that half the respondents are polled using one version of the question and half using the other version. If the number of subjects is sufficiently large, or if the subjects in the two groups are matched, the effect of the change in wording can be investigated.

A split poll can also be used to investigate 'interviewer' effects, such as whether the gender, age or status of the interviewer affects the answer given.

The next two activities contain ideas using split polls that you might wish to use or adapt in your own investigation.

Activity 2.18 Euthanasia

Two versions of the same question about euthanasia are given below. Carry out a survey to investigate whether the change of wording affects the answers that people give.

1 Terminally ill patients, in great pain, may wish to end their suffering. Mercy killing, or euthanasia, should be legalised to help them. Do you agree?

Both versions are leading questions

2 If mercy killing, or euthanasia, is legalised, terminally ill patients will feel under pressure to end their lives in order not to be a burden to their relatives and carers. Do you think euthanasia should be legalised?

The next activity is another example of a split poll survey.

Activity 2.19 Sexual discrimination

The same statement is used in both versions. The difference is that one version attributes this statement to Colonel Gaddafi, the other to Emmeline Pankhurst. Does this make any difference as to the proportion of people who agree with the statement? Are women more likely to agree in any case?

Ideally, you need over a hundred results for each of the two conditions in your experiment, in order to compare the results.

Version 1

Read the statement shown below and then tick one of the boxes in the scale.

'Discrimination between man and woman is a flagrant act of oppression without any justification.'
Colonel Gaddafi (Libyan Head of State).

Strongly agree	Agree	Undecided	Disagree	Strongly disagree
☐	☐	☐	☐	☐

Are you Male ☐ Female ☐

Version 2

Read the statement shown below and then tick one of the boxes in the scale.

'Discrimination between man and woman is a flagrant act of oppression without any justification.'
Emmeline Pankhurst (English Suffragette).

Strongly agree	Agree	Undecided	Disagree	Strongly disagree
☐	☐	☐	☐	☐

Are you Male ☐ Female ☐

Activity 2.20 Choosing your own investigation

1 a Choose a topic for your own investigation. Suggestions were given in Chapter 1 to give you some ideas, but you could choose to do something different. You should be ready to discuss your choice with your teacher at your next review.
b Prepare a brief and clear written statement of the aims of your investigation. This will help you to focus when you begin work on it. Don't try to do too much. Remember that your investigation is only intended to occupy 20 hours of your time.
c Identify the population of interest in your investigation, and consider how you might collect data to answer your research question. Decide whether your investigation will need a survey or an experiment to collect the data. Consider whether your research needs to focus on a subset of the whole population.
d Prepare some initial proposals about the design of the data collection. Before you collect any data you need to decide how you might analyse it, so you may need to modify your proposals after you have studied Chapters 3 and 4.

Reflecting on Chapter 2

What you should know

- the difference between a survey and an experiment
- different measurement scales and how to recognise them

- the meaning of the terms: simple random sample, systematic sample, stratified sample, quota sample, and cluster sample
- the advantages and disadvantages of different sampling methods
- the meaning of the terms: controlled, uncontrolled, explanatory and response variables
- things that can go wrong with experiments and some ways of dealing with them
- common mistakes in questionnaire design.

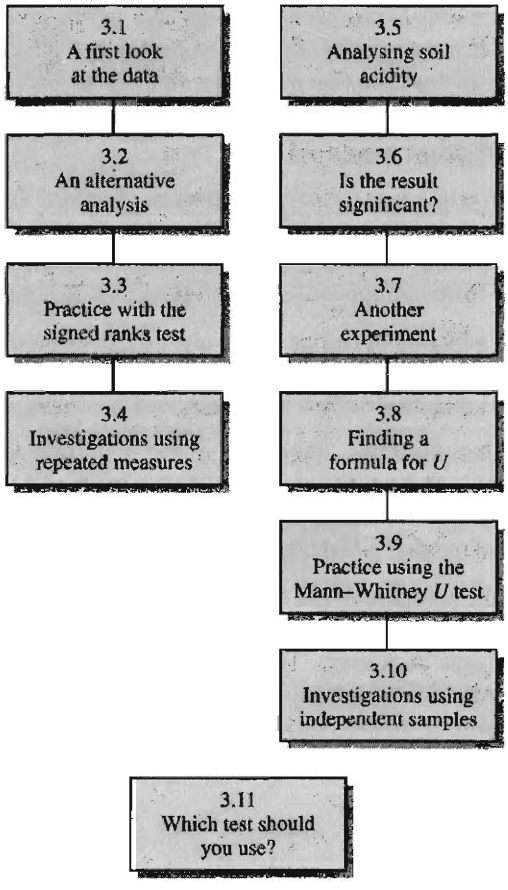
Preparing for your next review

- Be ready to discuss your plans for your own investigation in terms of the ideas discussed in this chapter.

There are no practice exercises for this chapter.

3 Looking for differences

This chapter builds on your work on hypothesis testing in the *Distributions* unit in Book 5. In that unit you used the sign test in order to examine the hypothesis that changes in performance were due to chance, against the alternative hypothesis that people improved with practice. In this chapter you will learn about two further tests which you can use to test whether two sets of data come from the same population or whether the populations differ in some way.



Activities 3.1 to 3.4 deal with the analysis of 'before and after' data: that is, data from investigations in which the same sample is measured on two occasions.

Activities 3.5 to 3.10 deal with the analysis of data collected from two independent samples.

Activity 3.11 is an exercise in deciding which is the appropriate test to use.

You can start either with Activities 3.1 to 3.4, or with Activities 3.5 to 3.10.

Once you have decided where to start, work the activities in sequence.

'Before and after' tests

In Chapter 2, you learned about three types of experiment: independent samples, matched groups and 'before and after' design. Each of these will give you two or more sets of data which you need to compare. When you analyse the results of these experiments, you need to test whether or not two sets of data can be said to have come from the same population.

In the *Distributions* unit in Book 5 you saw how the sign test can be used to analyse the results from matched groups and 'before and after' experiments. The memory experiment which follows is an example of a matched groups design; Activity 3.1 will remind you how to use the sign test to analyse the results.

A memory experiment

Do people find it easier to remember words or facts which make sense to them, as opposed to those which lack meaningful pattern? This idea has important applications for many forms of training, because it implies that information which has to be memorised should be presented in a meaningful way.

An experiment carried out to test this hypothesis used two lists of words. List A was made up of 24 familiar double-barrelled words. List B contained the same words, paired up randomly to give pairs which were less meaningful.

Each subject was asked to study list A for three minutes. The list was then taken away and the subject was asked to count backwards from 215 for one minute. The subjects were then asked to write down as many words as they could remember from the list; the number of pairs correctly remembered was recorded.

In the next part of the experiment, the subject was given list B and asked to study it for three minutes. The list was then removed and the subject asked to say the alphabet backwards. Then again they were asked to write down all the pairs of words which they could remember from this second list, and the number of correct responses was recorded.

Here are the two lists of words.

List A	List B
ICE CREAM	WASTE CREAM
INTER CITY	BRANDY ROAD
CUP FINAL	CAT BIN
ROLLING PIN	CHIMNEY BOWL
SCHOOL BUS	ICE SNAP
SMOKE SCREEN	STEERING FLAP
CAT FLAP	ROSE WHEEL
ROSE BOWL	CLOTHES FINAL
SHOE SHINE	SHOE TREK
CLOTHES LINE	CUP SHOOTER
PENCIL CASE	INTER SHINE
CUCKOO CLOCK	FAIRY BUS
CHIMNEY SWEEP	PEA POX
PONY TREK	PONY BARREL
BRANDY SNAP	PENCIL CART
WISHING WELL	SMOKE SWEEP
FAIRY TALE	WISHING LINE
CHICKEN POX	RING CLOCK
WASTE BIN	ROLLING WELL
RING ROAD	CUCKOO CITY
PEA SHOOTER	SCHOOL TALE
GO KART	DOUBLE PIN
DOUBLE BARREL	CHICKEN SCREEN
STEERING WHEEL	GO CASE

Table 3.1

Table 3.2 gives the results of the experiment.

Subject	List A	List B	Male/Female	Age	Occupation
1	9	0	M	36	head teacher
2	10	5	M	39	roofing contractor
3	10	6	F	22	telephonist
4	14	8	F	28	hairdresser
5	5	6	F	46	housewife
6	14	3	F	40	housewife
7	21	9	M	35	charge hand
8	5	5	F	28	civil servant
9	2	4	M	31	civil servant
10	12	13	F	38	secretary
11	11	3	M	28	civil servant
12	7	4	M	24	painter/decorator

Table 3.2

To begin with, you could investigate these results by using techniques which you already know from the units *Investigating data* in Book 2 and *Distributions* in Book 5.

Activity 3.1 A first look at the data

1 Display the results on a back-to-back stem-and-leaf diagram like this, and make a five-number summary for each list.

List A		List B
	0	
	0	
	1	
	1	
	2	

Figure 3.1

2 What do these analyses show? What can they not show?

3 For each person, calculate the difference between the number of pairs correctly remembered from list A and from list B.

4 Find the five-number summary for the differences in scores.

5 Comment on your answer to question **4**.

In the *Distributions* unit, you used the sign test to test for differences in 'before and after' experiments.

6 Write a null hypothesis and an alternative hypothesis for this experiment.

7 What assumptions are you making about the experiment, when you use the binomial distribution in your hypothesis test?

8 How many non-zero differences are there? How many of them are positive? How many are negative?



9 Use the binomial distribution with $p = 0.5$ to calculate the probability of getting this number or less of negative differences. Is the result sufficiently unusual for you to reject the null hypothesis?

You may have noticed that the three negative differences are all very small, -1 or -2, whereas the positive differences range in size from 3 to 12. The sign test is very simple, concentrating only on the sign of the differences. If you use it here, you are not making use of all the information available to you. In situations like this, it makes sense to think about the magnitude as well as the direction of the differences.

One test which does this is the **Wilcoxon signed ranks test**. For the memory test on double-barrelled words, the null hypothesis is that there is no difference between the scores on test A and test B. The alternative hypothesis is that the scores on test B are lower than on test A. Activity 3.2 shows how you can use the signed ranks test to test this hypothesis.

Activity 3.2 An alternative analysis

You will find it helpful to set out these calculations in a table.

1 **a** First calculate the differences between each score on list A and the equivalent score on list B. Some differences will be positive and some negative.
b Find the absolute value of each difference.
c Ignoring any zero differences, rank the absolute values from the smallest to the largest, keeping a note of whether originally the difference was positive or negative.

If several differences are the same, average their ranks. For example, the first two absolute differences are both 1. Their ranks are therefore 1 and 2, so you must give them both a rank of 1.5. Even if one difference is positive and one is negative, but they have the same absolute value, they must have tied ranks.

d Find the total of all the ranks awarded to positive differences.
e Find the total of the ranks awarded to negative differences.
f Call the lower of the two totals that you have calculated, T. As a check, work out the sum of all the ranks for non-zero differences, both positive and negative. Is this the same as the sum of your answers to **d** and **e**?

2 Now you are ready to test whether this is a significant result: that is, one which would be very unlikely to occur if the null hypothesis were true. In this case, you have to find the probability of obtaining a value for T of 6 or lower; this is the value you should have obtained in question **1**. You need to assume that each arrangement of pluses and minuses is equally likely.

a There are eleven ranks in all, each of which could be positive or negative. How many different equally likely outcomes are possible? In other words, how many different arrangements of pluses and minuses are possible?
b How many of these outcomes give a value for T of 6 or lower?

You can answer **b** by constructing a table showing possible ways of attaching + and – signs to the ranks to give low enough values of T. Because the null hypothesis is one-directional, you want only those for which the lowest total is for the negative ranks. The first few have been done for you.

Rank											T
1.5											
1	2	3	4	5	6	7	8	9	10	11	
+	+	+	+	+	+	+	+	+	+	+	0
–	+	+	+	+	+	+	+	+	+	+	1.5
+	–	+	+	+	+	+	+	+	+	+	1.5
–	–	+	+	+	+	+	+	+	+	+	3

Table 3.3

c Use your answers to **a** and **b** to calculate the probability that $T \leq 6$. Is the result sufficiently unlikely for you to reject the null hypothesis? What do you conclude?
d You used a one-tailed test because you predicted that scores on test B would be lower under the alternative hypothesis. If you are not sure in which direction the difference will be, you should use a two-tailed test. To do this, you also need to find

how many ways you can get values of T of 6 or lower for which the lowest total is for the positive ranks. How many of these will there be?

e Using a two-tailed test, what is the probability of obtaining $T \leq 6$? Is this a significant result?

f Do you think that taking the magnitude of the difference into account gives a fairer test of the null hypothesis for this experiment?

Summary – Wilcoxon signed ranks test

The word 'instrument' refers to any device for measuring the variable of interest to you. In many investigations this will be a test or a questionnaire.

You can use the Wilcoxon signed ranks test whenever you have two measurements for each subject, made using the same instrument, and you want to investigate the difference between these two measurements.

Assumptions

- The distribution of the differences between pairs of scores is approximately symmetrical.
- In these examples, the Wilcoxon signed ranks test is being used to test whether the median of the distribution of differences between pairs of scores is zero.

Method

- Calculate the difference between each pair of scores. Some will be positive and some negative.
- Ignoring the zero differences, rank the others, positive and negative, in terms of absolute value from the smallest to the largest.
- Find the total of all the ranks awarded to positive differences.
- Find the total of the ranks awarded to negative differences.
- The lower of these two totals is called T.
- To decide the significance of the result, you have to find the probability of obtaining a value of T equal to or lower than the calculated value.

Larger samples

Even for quite small samples, working out the probability takes quite a long time. With larger numbers, this task can become very difficult and time-consuming. Fortunately, there are tables which tell you the significant values of T for data sets of different sizes. You will find an excerpt from such a table in Appendix B, Table 4. As an example, you can use this table to test whether the result in Activity 3.2 is significant at the 5% level. You are using a one-tailed test with $n = 11$, so look in the table at the row with 11 in the first column, and the column headed 5% in the one-tail row.

One tail	10%	5%	2.5%	1%	0.5%
Two tails	20%	10%	5%	2%	1%
n					
11	18	(14)	11	7	5

Table 3.4

The lower tail critical value of T is 14, but the observed value of T from the experimental results is 6. Since the observed value of T is less than the critical value, you conclude that the result is significant at the 5% level.

Using computers

Better still, computer software, such as Minitab, can carry out the whole process for you. If you input the two sets of numbers, the computer will calculate the value of T and compute the probability of obtaining that value or less under the null hypothesis.

There are some things, however, which a computer cannot do for you. The computer cannot tell you whether or not it is legitimate to use the signed ranks test, or whether you should use a one-tailed test or a two-tailed test. Nor will the computer tell you what significance level is appropriate in particular circumstances. All these are decisions which you have to make for yourself.

Activity 3.3 Practice with the signed ranks test

For each of the three experiments described below, write down:

a a null hypothesis
b an alternative hypothesis
c whether you are using a one-tailed or a two-tailed test
d whether you use the Wilcoxon signed ranks test to test the significance of the results given in the question
e your conclusions.

1 The members of a randomly selected group of people are asked to weigh themselves immediately before going to bed, and then again first thing in the morning, wearing the same clothing on both occasions. Here are the weights they record.

Person	A	B	C	D	E	F	G	H	I	J	K	L
Mass (kg) in evening	28.0	35.0	54.5	68.0	70.5	42.5	51.0	55.0	50.0	63.6	69.1	31.0
Mass (kg) in morning	27.4	35.0	54.0	68.0	70.0	43.0	50.0	54.2	49.9	63.2	68.4	30.3

Table 3.5

2 Ten college physical education students are timed for swimming 100 m breast stroke and 100 m butterfly. Here are the results.

Student		A	B	C	D	E	F	G	H	I	J
Time in seconds	Breaststroke	92	80	102	105	81	96	88	76	100	97
	Butterfly	90	83	98	106	75	93	84	70	91	84

Table 3.6

3 Eight athletes run a 400 m race at approximately sea level, and at a later meeting run another 400 m race at high altitude. Here are their times in seconds.

Athlete		A	B	C	D	E	F	G	H
Time in seconds to run 100 m	Sea level	47.5	48.9	47.9	48.4	51.2	48.0	50.2	50.8
	High altitude	47.2	47.8	48.5	50.5	54.6	49.8	52.1	54.9

Table 3.7

Activity 3.4 Investigations using repeated measures

Here are some ideas for experiments that you can carry out yourself. Each of these experiments will give two measurements for each subject, so will enable you to use the Wilcoxon signed ranks test to analyse the results.

Some of these experiments have already been described in this section or in the *Distributions* unit in Book 5.

1 Word memorisation. Try the test described on page 34 on a group of volunteers.

2 Night and morning weights. Ask some people to weigh themselves night and morning, as described in Activity 3.3, question 1.

Why do you need to randomise the order?

3 Reaction times. Do people have different reaction times if they catch a reaction ruler with their writing hand as opposed to their non-writing hand? Use a coin or random numbers to randomise the order in which your subjects use their writing or non-writing hand.

4 Stereoscopic vision. When you look at an object, two pictures are sent to the brain, one from each eye. The fact that they are viewed from two slightly different positions enables the brain to judge distances and gives you three-dimensional vision. This is known as the stereoscopic effect.

This experiment is designed to test the strength of the stereoscopic effect. It will enable you to compare people's accuracy in judging distance when looking with only one eye, as compared to using both eyes.

You need a strip of paper about 1 m long and at least 20 cm wide. Divide it crossways into sections 2 cm wide. Shade in the centre section. This is the target. Number the sections, working outwards from the target, so that the target is 0, and the others are numbered from 1 upwards in both directions away from the target.

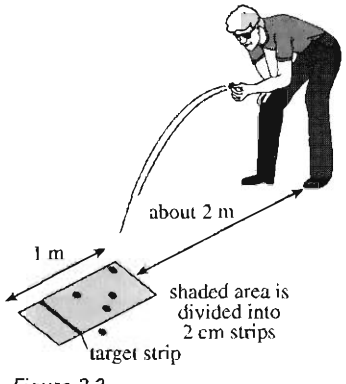


Figure 3.2

Ask your subject to stand about two metres from the target and throw a two-pence coin at the target. Allow each subject ten throws to practise. The score is the number of the section on which the coin lands. You will have to make a rule about what score to give when the coin lands on a line.

Use a coin or random numbers to randomise the order in which subjects attempt the test. Some subjects should aim for the target using both eyes for their first five throws and then using one eye only for the next five throws. Other subjects should do the test in the reverse order, using first one eye and then two. Allow your subjects to choose which eye they use for the 'one eye' condition.

For each subject, find the average score for the five throws using both eyes, and the average score for the five throws using one eye only. Is the score using both eyes lower than the score using one eye only?

Independent samples

You can use the Wilcoxon signed ranks test only when you have matched data: that is, when the same people have been measured on two occasions, or pairs of subjects have been chosen to match one another on other relevant variables. In many investigations this does not apply.

Suppose that you have data from two independently chosen samples and you are interested in finding out whether you can assume that they come from the same population, or whether they differ in some way. You can still use methods based on ranks.

The idea is best explained by an example.

Activity 3.5 Analysing soil acidity

As part of a study of soil acidity, a group of biology students carry out some fieldwork in the Peak District in Derbyshire. The students suspect that north-facing slopes will have more alkaline soils than south-facing slopes; this would be indicated by higher pH values in soil from the north-facing slopes. They collect samples of soil from both north- and south-facing slopes and measure the acidity of each sample. Here are their results.

Soil pH values (set A)	
North-facing slopes	South-facing slopes
6.5	5.5
7.5	6.0
7.0	6.75

Table 3.8

The students have to carry out a statistical analysis to decide whether the two sets of results are similar or different.

1 Write down the null hypothesis and the alternative hypothesis for this investigation.

2 a People will get a feeling for these results more easily if you present them in a visual form. To do this, plot the data on a number line, labelling each point as either N (if the sample comes from a north-facing slope) or S (if it comes from a south-facing slope). The graph will be particularly clear if you mark the N and S points in different colours.

You can now see at a glance the distinction between the two sets of measurements, but this graph still does not tell you how likely it is that you could have obtained results like this just by chance, even if there is no real difference between soils on north- and south-facing slopes.

b Here are two other sets of data. They are made-up, invented to show you some of the possible outcomes which you could get from an experiment like this. Draw separate number lines for each set, and plot the data in the same way as in part **a**.

Soil pH values (set B)	
North-facing slopes	South-facing slopes
8.0	4.0
7.0	5.5
6.5	5.0

Table 3.9

Soil pH values (set C)	
North-facing slopes	South-facing slopes
5.5	6.0
6.75	6.25
7.5	7.0

Table 3.10

3 For each of the data sets given in question **2**, you can calculate a statistic which measures the amount of separation of the two sets of results. Follow these steps to find this number.

a For each N result, count how many S results are less than it.



b For each S result, count how many N results are less than it.

c Now add the three numbers that you found in **a**. This gives you the number of S's under N's. Call this U_1. In the same way, add the numbers you found in **b**. This is the number of Ns under Ss. Call it U_2.

d Make up two or three more sets of data. Graph each set on a number line and calculate U_1 and U_2. In one set, have all the N's less than all the S's, but in the other sets have them mixed together.

4 How do think you can tell from the values of U_1 and U_2 whether there is a large separation between the N's and the S's or whether there is a lot of overlap?



The statistics U_1 and U_2 give you a way of measuring the distance, or separation, between two sets of numbers.

Do you think the amount of separation the students found in their results is sufficient evidence for them to decide that there is a difference in soil acidity between north- and south-facing slopes? To answer this, you need to think about probabilities.

As you learned in the *Distributions* unit in Book 5, it is always up to the person carrying out an investigation to decide how much risk of being wrong they are prepared to tolerate. A result which could be obtained by chance with a probability of 5% (0.05) or less, is regarded as improbable. In such a case you reject the null hypothesis. However, if the consequences of being wrong are very serious, you may want to be more cautious and use a lower probability level, such as 1% (0.01).

In the case of the soil acidity investigation, an error in the result would probably not be very serious, so the students are likely to decide that a probability of up to 0.05 is satisfactory.

In Activity 3.6 you will investigate a probability model which will help you to decide whether any result is significant: that is, unlikely enough for you to reject the null hypothesis.

Activity 3.6 Is the result significant?

The null hypothesis for the soil acidity investigation can be re-stated as:
'The soil pH values for north- and south-facing slopes do not form separate populations. Both samples come from the same population.'

You can reject the null hypothesis only if the values of U_1 and U_2 show that there is a high degree of separation: so high that it is very unlikely to have occurred by chance. Before you can decide how unlikely a particular value of U_1 or U_2 is, you need to work out the probabilities of the different possible outcomes.

1 Notice that any set of results can be represented by a sequence of three N's and three S's, in some order. The actual size of the measurements does not matter, only their order.

a Write an algorithm *FindU* which will generate randomly a sequence of three N's and three S's, so that any sequence is as likely as any other, and calculate the value of U_1 for this sequence. (This is a bit like tossing a coin until you get three heads or three tails.)

Implement your algorithm on your graphics calculator.

Why do you not also need to calculate U_2?

b Write a second algorithm, *ProbU*, to simulate the collection of many different sets of data from an investigation such as this. Your algorithm will need to call *FindU* repeatedly and record the frequency with which the different possible values of U_1 occur. It will need to deal with at least 100 sets of data.

c Use this simulation to estimate the probabilities of obtaining different values of U_1 and U_2.

2 A simulation can give you only approximate values of the probabilities. By thinking about the situation, and analysing it carefully, you can find them exactly.

a How many different possible arrangements of three N's and three S's are there?

b List all these arrangements. The first four are given in the table below.

c For each arrangement of N's and S's, calculate U_1 and U_2. The first four have been done for you in Table 3.11.

	Arrangement	U_1	U_2
1	SSSNNN	0	9
2	SSNSNN	1	8
3	SSNNSN	2	7
4	SNSSNN	2	7

Table 3.11

If you can, work with other people to find all the possible sequences and then share out the work of calculating U_1 and U_2.

3 Using the table that you produced in question 2, you can assign probabilities to each of the possible pairs of values of U_1 and U_2.

a If the null hypothesis is true, all the possible arrangements of N's and S's will be equally likely. What is the probability of each?

H **b** What is the probability of each possible value of U_1?

c How do the theoretical probabilities that you have calculated compare with the results of the simulation which you carried out in question 1?

Notice the symmetry in your table, with equivalent results at the top and the bottom. A result of $U_1 = 0$ and $U_2 = 9$ is equally as unlikely as a result of $U_1 = 9$ and $U_2 = 0$. Since the order in which the two samples are originally written – 'S and N' or 'N and S' – is quite arbitrary, there is no reason at all to discriminate between these two results. Therefore $U_1 = 0$ and $U_2 = 9$ must be taken together with $U_1 = 9$ and $U_2 = 0$, giving a total probability of 0.1. For this reason, the people who invented this test chose to focus on the lower of the two values U_1 and U_2. This lower value is referred to simply as U.

4 Use your table to list the possible values of U and the probability of obtaining each by chance.

The results collected by the students show quite a high degree of separation between the two samples, with only a slight overlap. From your table, you can see that the probability of obtaining the degree of separation shown by $U_1 = 8$ and by $U_2 = 1$ (that is, $U = 1$) is 0.1. However, $U = 0$ indicates even greater separation than $U = 1$, so if you reject the null hypothesis for $U = 1$ you also have to reject it for $U = 0$. This means that you must combine the probability of obtaining $U = 0$ with that for $U = 1$, and look at the probability that $U \leq 1$. This is equal to 0.2, which is much too high a probability for the null hypothesis to be rejected, and indicates that nothing particularly unusual has taken place.

As you can see, even the extreme results of $(0,9)$ or $(9,0)$ have a total probability of 0.1. So even this result would not have a low enough probability for the students to reject the null hypothesis.

This test was developed by statisticians H. B. Mann and D. R. Whitney in 1947 to analyse the degree of separation between sets of results like those you have been working with.

So what can be done? The difficulty is that the students have not collected enough results. Their samples are not large enough for them ever to be able to reach a decisive conclusion. They need to use larger samples.

The procedure you have used here is called the **Mann-Whitney *U* test.** It is a non-parametric test, because it does not make any assumptions about the underlying distribution.

Activity 3.7 Another experiment

1 A biology student carried out an experiment to see whether there is any difference between the percentage of brine shrimp eggs hatching under light and dark conditions. She expected that a higher percentage would hatch under light conditions. Her results for seven batches of eggs are shown below. One result, for dark conditions, was lost when a petri dish was broken. Luckily, there is no need to have equal numbers of results in the two sets in order to carry out the analysis.

Percentage of shrimp eggs hatching	
Light conditions	Dark conditions
80	53
66	61
57	65
61	

Table 3.12

a Write down the null hypothesis and the alternative hypothesis for this experiment.
b Show these results on a number line.
c Calculate the values of U_1 and U_2. (For tied results, count $\frac{1}{2}$ below.)
d On a number line, show a set of made-up results in which the two samples are completely separated, with all those kept under light conditions having a higher percentage hatching than those kept under dark conditions. Graph another set which has complete separation, but in the opposite direction. Calculate the values of U_1 and U_2 in each case.
e How many sequences of light and dark results are possible if you have four light and three dark? Assuming the null hypothesis, that all these sequences are equally likely, what is the probability of each?
f Assuming the null hypothesis, what is the probability of obtaining 0 as the value of U? Has the student collected enough data for there to be any possibility at all of a significant result, at the 0.05 level of significance? Give reasons for your answer.
g If the fourth petri dish for the samples hatched in the dark had not been broken, could a significant result have been possible?

2 **a** For an experiment with one sample of four and one of five results:

- what would be the values for U_1 and U_2 in the extreme case where the two samples are completely separated?
- what are the possible values of U?
- how many possible equally likely sequences are there? So what is the probability of each?
- what is the probability that $U = 0$?

Do not attempt to list all the possible sequences.

b Answer the same questions as in **a** for an experiment involving two samples with five results in each.

An easier method of calculating U

So far, you have worked with very small sets of data, so it has been possible, although tedious, to work out the probability of obtaining each of the possible values of U. However, as the size of your samples increases, this task very soon becomes impossible to handle.

For two samples with five results in each there are 252 different sequences. It would take some time to write down all of these, and even more time to calculate U_1 and U_2 for each. You will be relieved to know that there is a formula which you can use to calculate U_1 or U_2. There are also tables which show, for various sample sizes, which are the unusual values of U: that is, those with total probabilities of less than 0.05.

Activity 3.8 Finding a formula for U

Here again is the table of soil pH values.

Soil pH values (set A)	
North-facing slopes	South-facing slopes
6.5	5.5
7.5	6.0
7.0	6.75

Table 3.13

1 a Show both sets of results on a number line, but keep one sample on top of the line and one below. The first two have been done for you.

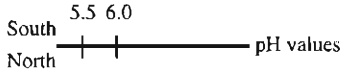


Figure 3.3

b Now rank both samples together, listing the first, second, third, and so on, setting out the ranks on a similar number line. Again, the first two have been done for you.

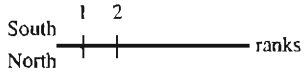


Figure 3.4

c Find the sum of the ranks for each sample. Call these R_1 and R_2.

d Let n_1 and n_2 be the numbers of results in sample 1 and sample 2. Show by substitution that the following formulae give the same values of U_1 and U_2 as those that you found in Activity 3.5.

$$U_1 = R_1 - \tfrac{1}{2}n_1(n_1 + 1)$$

$$U_2 = R_2 - \tfrac{1}{2}n_2(n_2 + 1)$$

2 To see where these formulae came from, let the results in sample 1 be $N_1, N_2, N_3, \ldots, N_{n_1}$, and let the results in sample 2 be $S_1, S_2, S_3, \ldots, S_{n_1}$, where both sets are arranged in order of size.

Choose any one of the N's, call it N_i, and let its rank in the combined sample be r_i.

a Explain why $r_i = i +$ the number of S's under N_i.

b Summing over all the N's gives:
$\sum r_i = \sum i +$ the total number of S's under N's. Use this to show that $U_1 = R_1 - \frac{1}{2}n_1(n_1 + 1)$.

c In a similar way, show that $U_2 = R_2 - \frac{1}{2}n_2(n_2 + 1)$.

Using tables

Calculating the probabilities of obtaining different values of U can be difficult and time-consuming. Luckily, these probabilities have been calculated and you can look them up in tables. The table in Appendix B at the back of the book (Table 5) gives the largest value of U which is significant for samples of different sizes: that is, for different values of n_1 and n_2. This is the upper-tail critical value of U. You calculate the lower-tail critical value of U by subtracting the value given in the table from n_1n_2. For you to reject the null hypothesis, the value of U calculated from your experimental data must be less than the lower–tail critical value.

Using a computer

Better still, you can use statistical software on a computer. If you input the two sets of data, the software will calculate the value of U for you and then tell you the probability of obtaining that value or smaller by chance on the assumption that the two samples came from the same population.

Find out how to use a statistical package to calculate U and find its significance. Test it on the soil acidity data and the shrimp egg data, and then use it in the activities that follow.

If you are using a statistical package such as Minitab for the Mann-Whitney U test, you may find that the output shows information about confidence intervals, but you should ignore this for now.

Summary – Mann-Whitney U test

The test assumes that

- a random sample is selected from each population
- the samples are taken independently of one another
- the distributions have approximately the same shape.

The null hypothesis is that the medians of the two populations are the same. The alternative hypothesis is that the medians are different.

The Mann-Whitney U test may be used for

- small samples
- any kind of numerical data, provided it is not nominal.

The test does not assume a normal distribution, as some other statistical tests do, so it is called a **non-parametric test**.

The formulae for calculating the test statistics are

$$U_1 = R_1 - \tfrac{1}{2}n_1(n_1 + 1)$$

$$U_2 = R_2 - \tfrac{1}{2}n_2(n_2 + 1)$$

where n_1 is the size of the first sample, n_2 the size of the second sample, R_1 the sum of the ranks of the first sample and R_2 the sum of the ranks of the second sample when the two samples are combined.

U is taken as the lower of the two results U_1 and U_2. To decide whether this value of U is sufficiently unusual to be significant, look at the table for the Mann-Whitney U test, using n_1 and n_2 to locate the required value. Your value of U must be lower than the value in the table in order to be significant.

A useful check on your calculations is that $U_1 + U_2 = n_1 n_2$.

Activity 3.9 Practice using the Mann-Whitney U test

Use either a computer or the formulae and tables to analyse the following sets of experimental data.

1 A mixed group of swimmers is randomly assigned to two groups. One group is given a target to aim for, while the other is not. Both groups are asked to swim 25 m breaststroke, and their times are recorded.

a Write down a null hypothesis and an alternative hypothesis.
b Does giving swimmers a target make any difference to their times?

	Swimming times in seconds for 25 m breaststroke									
No target	20.8	26.0	22.6	24.5	23.3	22.8	24.9	21.4	25.2	22.8
Target set	22.3	23.8	21.2	20.6	24.5	23.4	18.4	22.7	22.0	19.8

Table 3.14

c Can you suggest a better design for this experiment?

2 On the Derbyshire fieldwork trip, two other groups of students collect data on pH values for soils on north- and south-facing slopes. Here are the results for all three groups of students.

Soil pH values										
North-facing slopes	6.5	7.5	7.0	5.5	7.0	7.5	7.0	7.5	7.0	6.75
South-facing slopes	5.5	6.0	6.75	6.5	6.5	6.5	5.0	4.5	5.0	4.5

Table 3.15

Use the Mann-Whitney *U* test to test whether the two samples are different.

3 A nurse wonders whether mothers who smoke give birth to babies with lower birth weights. She asks all the women with children whom she knows whether or not they smoke and for the birth weights of their babies.

Table 3.16 shows her results.

Birth weights of babies in ounces according to whether or not the mother smokes															
Smoker	80	94	96	96	97	101	101	102	103	106	109	111	113	120	125
Non-smoker	108	110	112	119	121	122	122	123	124	127	128	130	143	157	

Table 3.16

a Is this a true experiment? Explain why.
b Comment on the selection of the sample. Is it likely to be representative of the local population?
c Which extraneous variables are not controlled?
d Write a null hypothesis and alternative hypothesis for this study.
e Use the Mann-Whitney *U* test to test for a difference in the two sets of results.
f Do you think that this study proves anything?

Activity 3.10 Investigations using independent samples

This activity contains ideas for questions which you might like to investigate. Since they all involve independent samples, you will be able to use the Mann-Whitney *U* test to analyse your results.

1 Are the reaction times of young people and older people different?

2 **a** Are the heights of males and females different?
b Are the weights of males and females different?
c Are the shoe sizes of males and females different?

3 Are people able to persevere at a simple task for longer if they are on their own, or if they are with another person?

Suggested task: sit on the edge of a bench or a chair. Extend one leg in a horizontal position for as long as possible.

Other tasks to try might be skipping, or bouncing a ball on the floor using the palm of the hand.

Use a coin to randomise the selection of subjects into individual or accompanied status for the experiment.

4 Some researchers have claimed that males are quicker than females at spatial tasks. Test this claim by designing an experiment to find out whether females are able to complete a maze more quickly or less quickly than males.

Alternatively, design an experiment to find out whether students taking mathematics and science are quicker at completing a maze than students studying humanities subjects.

Activity 3.11 Which test should you use?

For each of the experiments described below

a decide which experimental design has been used – either
'before and after' (repeated measures design) or
independent samples

b then select the correct statistical test to analyse the results – either
Wilcoxon signed ranks test, or
Mann-Whitney *U* test

c state a null hypothesis

d state an alternative hypothesis

e analyse the results using the statistical test that you have selected

f state your conclusions.

1 A university counsellor believes that hypnosis is more effective than the standard treatment given to students suffering from extreme anxiety in examinations. To test his belief, he randomly divides 22 students suffering from examination stress into two groups. One group receives the standard treatment and one the hypnosis treatment. When the treatments are concluded, each student is given an examination anxiety questionnaire. The scores of the two groups are then compared. High scores on the questionnaire indicate high anxiety.

Hypnosis treatment	25	24	34	41	24	43	49	32	23	45	30
Standard treatment	43	36	31	55	27	38	29	52	60	59	

Table 3.17

2 A psychology student has been told that men are better than women at abstract reasoning. The student is sceptical, so she decides to test this idea by randomly selecting a group of ten women and ten men from students at her college. She asks them to complete a test in abstract reasoning, and compares their scores.

Women	82	79	53	96	94	86	91	76
Men	69	87	59	93	81	66	75	95

Table 3.18

3 The personnel director of a large company wants to know whether it takes less time to type a standard monthly report using a word processor or an ordinary electric typewriter. A random sample of eight secretaries is selected from among the company offices. Each secretary is asked to type one report using a word processor and a similar one using an electric typewriter. The order in which they do this is

decided by tossing a coin. The times in minutes which they take to type the reports are recorded and compared.

Secretary	A	B	C	D	E	F	G
Word processor	15.3	16.4	15.8	20.1	14.8	15.3	17.0
Electric typewriter	18.0	16.1	16.5	22.3	13.9	16.2	18.1

Table 3.19

4 The director of an old people's home believes that exercise will improve the self-image of the residents in the home. To investigate this possibility he randomly selects 12 residents to participate in a specially designed exercise programme. Before they start, they are given a questionnaire to measure self-image. They are then enrolled in the programme and after two months their self-image is measured again and the two scores compared. The highest score possible on the test is 50, and high scores indicate high self-esteem.

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Before the programme	41	32	28	15	25	33	12	26	40	26	26	19
After the programme	39	36	33	17	31	43	10	38	44	34	32	26

Table 3.20

5 A cigarette manufacturer claims that the tar content of a new brand is lower than that of other low-tar cigarettes. A sample of this new brand is tested in a laboratory and compared to a random sample of another leading brand. The results give milligrams of tar for each cigarette tested.

New brand	15.3	15.8	16.2	15.1	14.9	15.3	15.7
Other brand	18.2	16.1	15.9	16.3	15.8	16.4	16.3

Table 3.21

6 A psychologist is interested in finding out whether left-handed and right-handed people differ in spatial ability. She randomly selects ten left-handers and ten right-handers from the students at her college, and gives them a test to measure spatial ability. She then compares their scores.

Left-handers	71	87	61	93	83	66	75	95
Right-handers	82	81	51	96	94	86	91	76

Table 3.22

Parametric and non-parametric tests

The main focus of this chapter has been on testing hypotheses about differences between populations. The tests which you have used, the Wilcoxon signed ranks test and the Mann-Whitney *U* test, are examples of what are called **non-parametric** or **distribution free** tests. The sign test, which you used in the *Distributions* unit in Book 5, is another example.

Non-parametric methods work with the ranks of the data instead of the values. This means that they are more resistant to error. This also means that you can use these tests with data measured on an ordinal scale. On the other hand, if your observations were measured on an interval or ratio scale, and the population can be modelled by a particular distribution, then parametric tests are usually more powerful. They require fewer items of data to establish a significant effect.

The non-parametric methods you have been using are also more **robust** than parametric tests: that is, they require fewer assumptions to be made about the data. Although, strictly speaking, these methods require data from a continuous distribution, they do not go far wrong when they are applied to discrete data, and in most cases they give a conservative result: that is, they tend to err on the side of caution. They are more likely to fail to find a significant difference that really exists than to find a spurious difference when, in fact, the two populations are the same.

Parametric methods assume that the observations to be tested have been drawn from a normal distribution. If the underlying distribution is close to normal, these methods do not go very far wrong, but it can often be difficult to judge whether this is the case. They are called parametric because they are concerned with estimating the parameters of the normal distribution, the mean and the variance. But they are very sensitive to a few extreme or outlying observations, so just a few errors in the data can make a big difference to the outcome. The mathematics underlying parametric methods is difficult and is usually omitted from elementary texts. This means that students learn techniques without knowing the explanations for them.

If you would like to find out more about the equivalent parametric methods, you can look up a test called the t test in any standard statistics textbook. Your teacher has a list of suggested books.

Reflecting on Chapter 3

What you should know

- the difference between non-parametric and parametric tests
- how and when to use the Wilcoxon signed ranks test to test for differences in an experiment using repeated measures
- how and when to use the Mann-Whitney *U* test to test for differences in an experiment with two independent samples
- how to use tables for these tests.

Preparing for your next review

- Be ready to explain how you would decide which is the appropriate test to use in an investigation.

- Think about how you might analyse the data in the investigation which you are planning to carry out as part of your work in this unit.
- Answer the following check questions.

1 An office manager wants to compare the effectiveness of two training courses for typists. Seven typists are sent on training courses, three on one course and four on the other. The improvements in their typing speeds in words per minute are shown in Table 3.23.

Course	A	10	12	9	15
Course	B	4	7	14	

Table 3.23

Explain how to calculate the Mann–Whitney U statistic to measure the separation between these two samples, and how to calculate the probabilities of obtaining different values of the statistic. Refer to Activities 3.5 and 3.6 if you need to.

2 Two systems of calf-rearing are compared on five pairs of identical twins, and the calf weight in kg of each was recorded at the end of the programme.

	Calf pairs				
	1	2	3	4	5
System A	107	111	103	117	153
System B	115	120	100	125	169

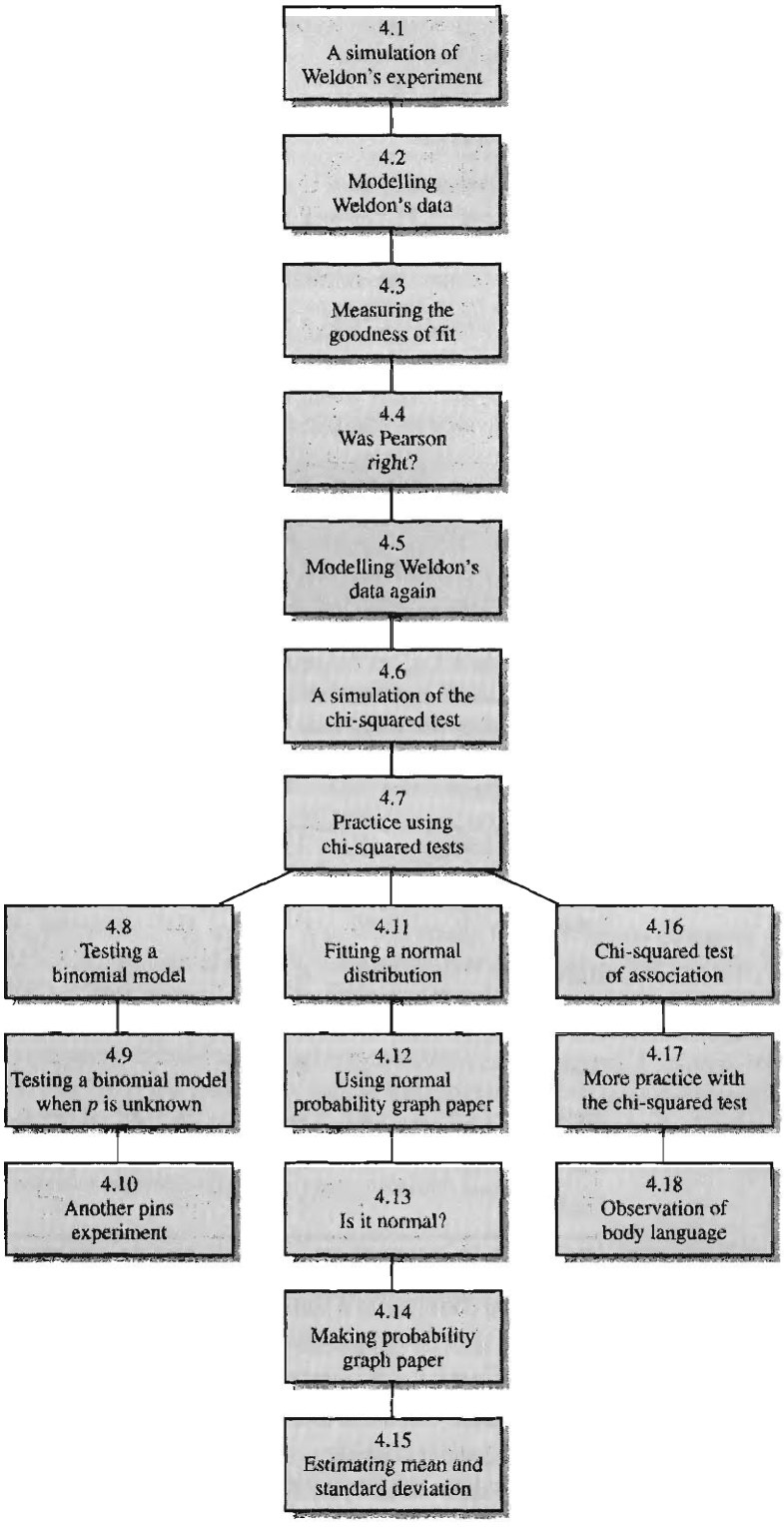
Table 3.24

Explain briefly how to calculate the signed rank statistic T for use in the Wilcoxon signed ranks test, and how to calculate the probability distribution of T in this case.

Practice exercises for this chapter are on page 101.

4 Fitting a model to the data

In this chapter you will learn about the χ^2 distribution as a way of testing whether some real data fits a hypothesised model.



In Activities 4.1 to 4.7 you will learn a way to test how well a theoretical model fits the observed data. These activities should be worked in the order given.

The chi-squared model is introduced to test the fit of a binomial model to some experimental data.

Activities 4.8 to 4.10 contain more examples of fitting a binomial distribution to a set of data.

Activities 4.11 to 4.15 are about fitting a normal distribution to a set of data. In Activity 4.11 you will test the fit of a normal distribution using a chi-squared test. In Activities 4.12 to 4.15 you will learn how to use normal probability graph paper to see whether the normal distribution is a suitable model.

In Activities 4.16 to 4.18 you will learn how to use the chi-squared test to find whether two nominal variables, such as hair colour and gender, are associated. Activity 4.18 is a suggestion for a longer study which you might have chosen to use for your own statistical investigation.

Activities 4.1 to 4.7 should be tackled first, and in order.

Activities 4.8 to 4.10, 4.11 to 4.15 and 4.16 to 4.18 form three separate groups. You can work on any of them after you have completed Activity 4.7. Within each group you should work on the activities in order.

Table 4.1 gives the results of a drawing pins experiment, generated by a group of students. Each student drops ten pins about 50 times, and records the number of pins landing point-up. One student's results look suspiciously different from the others. How can you know whether he has done something different, or whether this is just chance?

Number of pins landing point-up out of ten	1	2	3	4	5	6	7	8	9	10	Total number of throws
Student A	0	0	0	3	3	8	15	18	2	1	50
Student B	0	0	0	1	7	8	14	13	6	1	50
Student C	0	0	0	0	10	8	9	17	5	1	50
Student D	0	0	0	1	5	4	12	10	3	0	35
Student E	0	1	0	1	12	9	12	9	6	0	50
Student F	0	0	3	9	15	16	23	20	11	3	100
Student G	1	2	11	9	6	7	8	2	3	1	50

Table 4.1

You could build a simulation and explore how often such unusual results occur. You will have an opportunity to look more closely at this set of data in a later activity.

A medical researcher, who is trying to assess the effectiveness of a new measles vaccine, conducts an experiment on a group of 200 children. The children are allocated at random into two groups of 100. The children in one group receive the vaccine, and the children in the other group receive only the inert diluting agent in which the vaccine is normally administered. After one year the numbers of children who have contracted measles in both groups are counted; the results are shown in Table 4.2.

	Caught measles	Did not catch measles
Vaccinated group	10	90
Control group	34	66

Table 4.2

How can you judge whether the new vaccine is effective, or whether the apparent reduction in the rate of infection is only due to chance? Again, you could build a simulation, assuming a certain probability of getting measles, and see how often such apparent reduction occurs.

In the *Distributions* unit in Book 5, you learned how to use the binomial and normal distributions as possible models for real data. You can explain some variation by assuming that you are looking at a sample drawn from a binomial or a normal population. So far, the only way you have for telling whether you can explain the variation by chance in this way is to run a computer or calculator simulation of the model and see how much variation there is in each of several hundred or even thousands of cases. This gives you some feel for how chance variation behaves, but it takes a long time to build up evidence of this kind, even with the help of a computer.

Estimate how long this might have taken.

In 1894, an eminent zoologist, Professor W. F. R. Weldon, FRS, was exploring the laws of inheritance and needed to know whether the variation in his results might be attributed to chance. He did not have access to the technology that you have used in this course to simulate randomness when he needed to collect sufficient data. In order to collect enough results from fair experiments following a binomial distribution, he asked various people, including his wife, to record the outcomes of 26 306 throws of 12 dice.

He asked one clerk in the office of University College to toss groups of 12 dice and record the number of dice showing either five or six. The clerk did this 7006 times! Weldon later showed his data, including the clerk's results, to the statistician Karl Pearson, who concluded that the clerk's data contained errors, because the frequencies differed so much from the theoretically expected frequencies. Pearson therefore rejected the clerk's data as unacceptable. Weldon was upset about Pearson's conclusion as he had always found his clerk very reliable. Bear in mind as you go through the following activities how easy it would be in this experiment to confuse a dice showing four spots with one showing five.

Activity 4.1 A simulation of Weldon's experiment

To simulate the experiment carried out by Weldon's clerk, you really need a powerful computer with a spreadsheet or a statistical package, but you can quite easily simulate on a graphics calculator the experiment for 700 throws of 12 dice.



1 Weldon's data are shown in Table 4.3. Produce a bar graph of these results, and use the results to estimate the proportion of the clerk's dice throws that showed five or six.

Number of 5s or 6s	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
Observed frequency	45	327	886	1475	1571	1404	787	367	112	29	2	1	0	7006

Table 4.3

2 If the clerk was accurate, and the dice were not biased, what proportion of individual dice throws would you expect to score a five or a six?

Write a null hypothesis for the distribution of the number of fives or sixes.

Write an alternative hypothesis.

3 Either write a program for your graphics calculator to simulate tossing 12 dice 700 times and recording the number of fives or sixes in each trial, or write a computer simulation of Weldon's experiment for 7006 trials.

4 Draw a bar graph of the results of your simulation on a scale comparable with the one which you used to graph Weldon's data, and compare the two graphs. Does the result of your simulation appear to be consistent with Weldon's data? Justify your answer.

It is a good idea to use a small number of trials such as 20 while you are writing and testing your program, since it will take several minutes to run with 700 trials.

An algorithm for a graphics calculator program is given in the Hints if you need it.

To test for unusual results there is a quicker way than simulating lots more results for comparison. The frequencies obtained in the original experiment are called the **observed** frequencies. The **expected** frequencies are those which, theoretically, would be expected in an ideal situation. Obviously, any real results will deviate slightly from these expected frequencies. Each repetition of the experiment using unbiased dice will generate a different set of frequencies, but for a large number of repetitions the deviation of the observed frequencies from the expected frequencies should be slight.

But how much variation is acceptable? How large will the differences have to be before they can be regarded as evidence that the observed data do not fit the model? This was the question which interested Weldon.

In Activity 4.2, you will use the binomial distribution to model the distribution of the number of fives and sixes.

Activity 4.2 Modelling Weldon's data

1 Weldon modelled his data using a binomial distribution. List the assumptions which have to be made to use this model, and give a justification for each assumption.

2 If you use a binomial model, what are the values of n and p? Check your answer against the null hypothesis and alternative hypothesis that you gave as answers to question **2** of Activity 4.1.

The iterative method for calculating binomial probabilities is described in the *Distributions* unit in Book 5.

H 3 Calculate the probabilities for the number of fives and sixes in 12 dice throws. It is best to use an iterative method to do this, on a spreadsheet or a graphics calculator. Notice that a useful check is that the total of the probabilities should be one.

4 Use the theoretical probabilities to calculate the number of times out of 7006 rolls of 12 dice that you would expect to get each number of fives and sixes. This will give you the expected frequencies for each possible number of successes. Check that the total of the expected frequencies is correct.

H

You now have a set of expected frequencies which you can compare with the observed frequencies given in Activity 4.1. In Activity 4.3 you will develop a way of measuring how well the model fits the data. This is often called measuring the **goodness of fit**.

Activity 4.3 Measuring the goodness of fit

1 On the same scale as you used for Weldon's observed data, draw a bar graph of the theoretically expected frequencies. Compare the differences.

The difference between the observed frequency and the expected frequency is a measure of how well the data fit the model.

2 Make and complete a table similar to Table 4.4.

Number of dice cast with 5 or 6 points (x)	O Observed frequency	E Expected frequency	$O-E$ Deviation
0	45		
1	327		
2	886		
3	1475		
4	1571		
5	1404		
6	787		
7	367		
8	112		
9	29		
10	2		
11	1		
12	0		

Table 4.4

3 Calculate the total for $(O-E)$. What do you notice? Will this always be true? Explain your answer.

4 Remembering how you calculate variance, suggest a way of overcoming this. Explain why this works.

5 Notice that the deviation for $x = 0$ is about the same as the difference for $x = 3$. Why is it not sensible that they both contribute equally to the measure of goodness of fit? Suggest reasons why dividing by E might be helpful. What problem can you see with dividing by E?

6 If the expected frequency is very small, then the relative error is more likely to be large than if the expected frequency is large, even if the model fits quite well. Explain why this is so.

7 To overcome the problem identified in question **6**, it is usual to combine groups which have a small expected frequency with the adjacent groups. In this case you need to combine the last three groups together. When you have done this, calculate $\sum \frac{(O-E)^2}{E}$.

Which results are unusual?

In order to judge whether this result is unusual, you need to know the distribution of $X^2 = \sum \frac{(O-E)^2}{E}$. The distribution of $\sum \frac{(O-E)^2}{E}$ is very close to the theoretical distribution known as χ^2 (or chi-squared). In fact, there is a family of χ^2 distributions, each with a different shape depending on the number of **degrees of freedom** denoted by ν. For any χ^2 distribution, the number of degrees of freedom is the number of independent free choices which can be made in allocating values to the expected frequencies.

χ is the Greek letter chi (pronounced 'kie', to rhyme with 'tie' and 'dye') equivalent to the letter 'x'.

ν is pronounced 'new' and is the Greek letter equivalent to 'n'.

In this text, X^2 always refers to the value calculated from the data using the formula, and is distinguished from the theoretical distribution chi-squared which it follows approximately.

The shape of the χ^2 distribution

The shape of the χ^2 distribution is different for each value of ν, and the equation of the distribution is beyond the scope of this book. The mean of each of these χ^2 distributions, though, is always the number of degrees of freedom ν. The distribution is positively skewed, becoming increasingly symmetrical as ν increases.

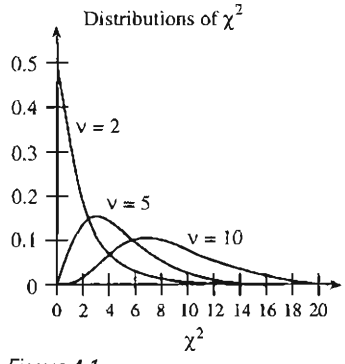


Figure 4.1

The χ^2 distribution for some values of ν

Activity 4.4 Was Pearson right?

1 Look back to the data in Activity 4.3 and imagine that you have to invent the observed frequencies. After you have combined the last three groups together, eleven numbers are required, and the total of all eleven must be 7006. In how many of them would you have a choice about the value?

Your answer to question **1** is the number of degrees of freedom, ν. In general the value of ν is given by ν = number of classes or groups – number of constraints.

2 Use Figure 4.1 to sketch the χ^2 distribution for $\nu = 10$, and mark on your sketch the 5% tails of the distribution given in Table 3 in Appendix B.

3 What value of X^2 was generated by Weldon's biased dice experiment?

4 A very low value of X^2 implies that the values of the observed and expected frequencies are very close. A high value of X^2 implies a poor fit between the observed and expected frequencies. When you use a χ^2 test of how well a model fits the data, you will not usually be concerned if the observed and expected frequencies are very close. You will therefore usually choose a one-tailed test for goodness of fit, using the χ^2 distribution. The model is rejected if the value of X^2 generated by the data is greater than the critical value.

Refer to this test to write your conclusions about whether Pearson was right to reject the data generated by Weldon's clerk.

Summary

- The χ^2 goodness of fit statistic is calculated as $\sum \frac{(O-E)^2}{E}$.
- The number of degrees of freedom, ν, is given by
  ν = the number of classes – the number of constraints.
- In situations where you are fitting a dataset to a uniform distribution, the only constraint on the expected frequencies is the total frequency.

Karl Pearson rejected Weldon's results as unreliable, but is it fair to blame the clerk? Both Pearson and Weldon were aware of the possibility that it might be the dice that were biased, rather than the clerk being mistaken. If the dice were biased, then the model developed in Activity 4.2 would not be appropriate, and a different approach is required. In Activity 4.5 you will develop a model so that you can test the reliability of the clerk when the probability of each dice showing five or six is unknown.

Activity 4.5 Modelling Weldon's data again

1 If the dice are biased and the clerk is reliable, would you still expect the data to fit a binomial distribution? Check whether the conditions still apply.

2 Write a null hypothesis for this situation, and then write an alternative hypothesis.



3 The problem now is that you don't know the value of p, the probability of getting a five or a six on each of the 12 dice, so you need to estimate it from the data. What is your best estimate of p?

4 Use this estimate of p to calculate the probabilities for the number of fives and sixes in twelve dice. Use the iterative procedure, as you did in Activity 4.2.

5 Use these theoretical probabilities to calculate the expected frequencies of different numbers of fives and sixes out of 12 dice.

6 Calculate the X^2 statistic $\sum \frac{(O-E)^2}{E}$ for this expected frequency distribution against the observed frequencies recorded by Weldon's clerk.



7 How many degrees of freedom are there this time?

8 Is the value of $\sum \frac{(O-E)^2}{E}$ so large that you would reject the null hypothesis?

9 There are possible sources of error in the data from Weldon's clerk which might look rather like biased dice. Suppose that the clerk had made the occasional mistake of recording a dice score of four as five. This might very easily happen in the course of 7006 trials! Explain why the tests you have done in Activity 4.3 and in this activity will not show this possibility.



10 Suggest some other possible sources of error in Weldon's data. Devise a way to analyse one of these sources of error.

During your investigation of the data from Weldon's experiment, you were not able to run a simulation of the experiment many times because the simulation takes so long to run. Therefore, you have not yet been able to generate a frequency distribution for $X^2 = \sum \frac{(O-E)^2}{E}$. In Activity 4.6 you will do this for an experiment which is easier than Weldon's to model on a graphics calculator. In it you will explore how the χ^2 test can go wrong.

Activity 4.6 A simulation of the chi-squared test

If you suspect a dice is biased, how can you tell?

The scores in Table 4. 5 were obtained using a dice suspected of being biased. To score one only seven times out of 120 throws seems to be a very low result.

Score	1	2	3	4	5	6
Frequency	7	14	18	23	26	32

Table 4.5

1 Produce a bar graph of these results, and calculate the mean score. Briefly record what you find.

2 If the dice is not biased, how many times would you expect to score each number in the long run? Write a null hypothesis for the distribution of scores. Write an alternative hypothesis.

3 Calculate the value of $X^2 = \sum \frac{(O-E)^2}{E}$ for the results given at the start of this activity, using the expected frequencies that you found in answer to question **2**. Explain in your own words what the χ^2 test tells you about the chance of observing such a set of results from a fair dice.

4 For a statistical package or a spreadsheet, or for a graphics calculator , write a program, *DiceFreq*, to simulate rolling a dice 120 times and recording the frequency of each score as you go along.

5 Use Table 4.6 to help you to calculate the value of $X^2 = \sum \frac{(O-E)^2}{E}$ for the results of your simulation.

Score	O Observed frequency	E Expected frequency	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
1		20			
2		20			
3		20			
4		20			
5		20			
6		20			

Table 4.6

6 Write a program using *DiceFreq* to calculate and record $\sum \frac{(O-E)^2}{E}$ for 100 simulations. Run it for five simulations first until you are sure that it works. Record your results in a copy of Table 4.7. Produce a histogram to show the frequencies of the values of $\sum \frac{(O-E)^2}{E}$, and comment on the shape of the distribution.

If you use a graphics calculator for question 6, make sure that you test your program thoroughly for a small number of simulations before trying to run for all 100. The full 100 will take about twenty minutes to run.

$X^2 = \sum \frac{(O-E)^2}{E}$	Tally	Frequency
$0 \le X^2 < 2$		
$2 \le X^2 < 4$		
$4 \le X^2 < 6$		
$6 \le X^2 < 8$		
$8 \le X^2 < 10$		
$10 \le X^2 < 12$		
$12 \le X^2 < 14$		
$14 \le X^2 < 16$		
$16 \le X^2$		

Table 4.7

7 Use the graphics calculator programs, or computer simulations, to explore hypothesis testing using the χ^2 test. Using a statistical package or a spreadsheet, the simulations are easy to do by adapting your earlier programs, but they are very slow on a graphics calculator. How often does the χ^2 test give a significant result when you use it to test 12 fair dice? Predict how often you think this will happen, then use your simulation programs to test your prediction.

8 Now try changing the theoretical frequencies of the dice to create a bias. For example, you might simulate a dice for which the probabilities of the different faces are given in Table 4.8.

Score	1	2	3	4	5	6
Frequency	0.1	0.1	0.2	0.2	0.2	0.2

Table 4.8

Use your simulation to estimate the probability that a χ^2 test using this dice would conclude that it is not biased. How often does your test give a significant result?

You have seen how to use the χ^2 statistic to test the fit of a binomial model when you know the value of p, and when you don't know the value of p. In the next three activities, you will practise using the χ^2 statistic to test the goodness of fit of different theoretical models of expected frequencies against the data observed in experiments.

Activity 4.7 Practice using chi-squared tests

For each experiment state:

- a null hypothesis
- an alternative hypothesis

This degree of detail is required for a statistical report so that its readers will be able to check your conclusions by reproducing your experiment if they wish.

- ν, the number of degrees of freedom for a χ^2 test
- the expected frequency for each category
- the value of $\sum \frac{(O-E)^2}{E}$ for the results given
- your conclusions.

1 An opinion poll is carried out to ask students which brand of cola drink they prefer. They are asked to choose from two leading names and a supermarket brand. Do the results, which are in Table 4.9, show any significant preference?

Leading brand 1	Leading brand 2	Supermarket own brand
80	55	15

Table 4.9

2 A researcher is interested to find out whether people think that fat people are more jolly than thin people. Eighty people in a random sample are asked: 'Do you think that fat people are more jolly than thin people?' The results are in Table 4.10.

Yes	No
47	33

Table 4.10

Is this significant?

3 The head of the marketing division of a leading soap manufacturer must decide between four different wrappers for a new brand of soap. She decides to run an experiment by having soap in all four wrappers placed for sale in each of five supermarkets. Table 4.11 shows the amount sold in the course of two weeks.

Wrapper	A	B	C	D
Number Sold	95	98	130	82

Table 4.11

Is there sufficient evidence to show a preference among consumers? If so, for which wrapper should she opt?

4 A pure-breeding pink-flowered tall variety of *Lathyrus* (sweet pea) is crossed with a pure-breeding white-flowered dwarf variety. The resulting seeds are sown and found to produce tall pink-flowered plants (these genes are dominant, but the other characteristics are carried by the plants as non-dominant genes). These are then cross pollinated with dwarf white plants and the seeds sown. The numbers of plants at this stage of the experiment are shown in Table 4.12.

Tall pink	Tall white	Dwarf pink	Dwarf white
80	33	39	8

Table 4.12

A botanist tells you that these four types should occur in the ratios 9:3:3:1. Use this as the null hypothesis. Are the actual results significantly different from this prediction?

Activity 4.8 Testing a binomial model

Two sixth formers conduct a survey of fun-size Smarties to see whether the distribution of red Smarties in fun-size boxes follows a binomial distribution. In fact, the number of Smarties in a box may vary, but for the purpose of constructing a model, you can assume that there are 16 Smarties in each fun-size box. Table 4.13 shows the frequency distribution for the number of red Smarties in a total of 40 boxes.

Number of red Smarties	Frequency
0	11
1	7
2	10
3	5
4	4
5	3
6 or more	0

Table 4.13

1 Draw a bar-line graph for these results, and calculate the mean and variance.

H 2 If you use a binomial model, what are the values for n and p? Write a null hypothesis and an alternative hypothesis.

3 Calculate the probabilities for the numbers of red Smarties using a binomial model. It is often helpful to have an iterative method to generate probabilities for a binomial model. You saw this method in the *Distributions* unit in Book 5.

Use a spreadsheet or statistical package if possible for questions 3, 4 and 5.

H 4 Use your theoretical probabilities from question 3 to calculate, for 40 boxes, the number of times that you would expect to see a box with x red Smarties. This will give you the expected frequencies for the values of x given in the table for probabilities in question 3.

5 Next calculate the value of $\sum \frac{(O-E)^2}{E}$. As very small values of E will artificially inflate the value of $\sum \frac{(O-E)^2}{E}$, expected frequencies below five are not used and are usually combined with an adjacent group to give a higher value of E.

The minimum of 5 for E is just a cautious 'rule of thumb' to ensure that the distribution of $X^2 = \sum \frac{(O-E)^2}{E}$ is approximately χ^2, but you should adhere to it as closely as possible, without losing too many values from the table of results.

6 How many classes are there in this table? How many constraints? What is the value of ν?

7 Use tables to find the upper 5% value of χ^2_4. Is the result obtained for this survey significant?



8 Can you reject the null hypothesis? What is your conclusion regarding the distribution of red Smarties in fun-size boxes?

Activity 4.9 Testing a binomial model when p is unknown.

A group of students conduct an experiment in which ten drawing pins are dropped one at a time on to the floor. For each group of ten pins, the students count how many pins land point-up and record this result.

They repeat this until the the class has results for 200 sets of ten pins. Table 4.14 shows their results.

Number of pins point-up	Frequency
0	11
1	39
2	47
3	53
4	46
5	4
6 or more	0

Table 4.14

The students have no idea, before they start the experiment, of the probability of a pin falling point-up. So the value of p must be estimated from the experimental results, if a binomial distribution is to be used as a model.

1 Use the frequency table to calculate the mean number of pins landing point-up.

2 Use this value of the mean to find p. Remember that, for the binomial distribution, $\mu = n \times p$. Use this expression rearranged to find μ, given that $n = 10$ (since drawing pins are dropped in groups of ten).

3 Write a null hypothesis and an alternative hypothesis.

4 Calculate binomial probabilities for $x = 0, 1, 2, 3, 4, 5, 6$, or more.

5 Use these binomial probabilities to find the expected frequencies, by multiplying each probability by the total frequency, 200.

In this situation, there are two constraints on the expected frequencies:

a the total frequency must equal 200

b the total for $\sum fx$ must equal 496. This second constraint follows from the fact that the mean must take the value found in this particular experiment to give the value of p estimated from the data.



6 a How many of the expected frequencies are less than five?

b Will you need to amalgamate any of the classes?

c How many classes can you use?

d How many degrees of freedom are there?

7 Find the value of $\sum \frac{(O-E)^2}{E}$.

8 Use tables to find the value of χ^2 exceeded only in 5% of genuine experiments.

9 Is your result significant?



10 What are your conclusions?

Activity 4.10 Another pins experiment

You are now in a position to look again at the data from the start of the chapter and to test any significant difference between the odd looking data generated by student G, in Table 4.15, and the rest of the data.

Number of pins landing point-up out of ten	1	2	3	4	5	6	7	8	9	10	Total number of throws
Student A	0	0	0	3	3	8	15	18	2	1	50
Student B	0	0	0	1	7	8	14	13	6	1	50
Student C	0	0	0	0	10	8	9	17	5	1	50
Student D	0	0	0	1	5	4	12	10	3	0	35
Student E	0	1	0	1	12	9	12	9	6	0	50
Student F	0	0	3	9	15	16	23	20	11	3	100
Student G	1	2	11	9	6	7	8	2	3	1	50

Table 4.15

1 Calculate the mean number of pins landing point-up out of ten for each student. Comment on what you see.

2 State a null hypothesis and an alternative hypothesis.

3 Investigate the fit of a binomial distribution to the frequencies given in Table 4.15



a including student G

b excluding student G.

4 If you can find a distribution to fit the results excluding student G, test against this distribution the fit of the data from student G only. Write a paragraph to summarise your conclusions.

Checklist for deciding whether a binomial distribution is a suitable model

1 Is the data set discrete?
2 Draw a graph. Is the shape of the graph uni-modal?
3 Is there a fixed number of trials, n?
4 Can results be categorised into two groups: 'success' and 'failure'? Is any particular result independent of earlier results?
5 Is there a theoretical value for p (the probability of 'success')? If not, p can be estimated by using the mean of the data results:

$$\text{mean} = n \times p \text{ so } p = \frac{\text{mean}}{n}.$$

6 Is it sensible to assume that the probability of success is the same for every trial?
7 For a test, where p is already known
ν = number of classes $- 1$;

where p is not known in advance
ν = number of classes $- 2$.

Testing the suitability of a normal model

It is often important to know whether the data set fits a normal distribution, since, if it does, there are many statistical methods, such as the parametric methods that were discussed briefly in Chapter 3, which you can use.

Activity 4.11 *Fitting a normal distribution*

The set of reaction times in Table 4.16 was obtained by using a computer, rather than a reaction ruler. Could the normal distribution be used as a suitable model for this set of data?

Reaction time in hundredths of a second	Frequency	True interval	Mid-value x
10–11	2	9.5 – 11.5	10.5
12–13	4	11.5 – 13.5	12.5
14–15	3	13.5 – 15.5	14.5
16–17	5	15.5 – 17.5	16.5
18–19	8	17.5 – 19.5	18.5
20–21	12	19.5 – 21.5	20.5
22–23	7	21.5 – 23.5	22.5
24–25	4	23.5 – 25.5	24.5
26–27	4	25.5 – 27.5	26.5
28–29	1	27.5 – 29.5	28.5

Table 4.16

1 Draw a histogram to illustrate this frequency table.

Before you can find any normal probabilities, you will have to estimate the mean and standard deviation from the data. This will place three constraints on the expected frequencies

- the total frequency must equal 50
- the mean must be estimated from the data, fixing the value for $\sum fx$ for the expected frequencies
- the standard deviation must be estimated from the data, and this fixes the value of $\sum fx^2$ for the expected frequencies.

2 Use your calculator or a spreadsheet to find

a the mean

b the standard deviation.

3 The stages **a**, **b**, **c**, and **d** are demonstrated for you in the table overleaf.

a Remember that the standard normal distribution uses units corresponding to the standard deviation measured from the mean, so you need to convert each interval boundary, X, into a Z value using $Z = \frac{X - \mu}{\sigma}$. Look at the *Distributions* unit in Book 5 if you a need a reminder of this. You can program your graphics calculator to perform the same sequence on the upper value of each interval. The task is even easier on a spreadsheet.

b Look up your values on normal distribution tables and write down the area to the left of each. Some spreadsheets have a defined function to do this for you.

c These are cumulative probabilities, as tables always give the area to the left for positive values of Z. Convert these to probabilities by subtracting each one from the next one lower down.

d Use these probabilities to calculate the expected frequencies. Remember that the total frequency should be 50.

These stages of working are best set out in a table such as Table 4.17.

Upper limit	a Z		b Area to left	c Probability of this interval	d Expected frequency
9.5	$Z=\frac{9.5-19.62}{4.33}$	$=-2.34$	0.00964	0.0096	0.5
11.5	$Z=\frac{11.5-19.62}{4.33}$	$=-1.88$	0.0301	0.0205	1.0
13.5		-1.41	0.0793	0.0492	2.5
15.5		-0.95			
17.5					
19.5					
21.5					
23.5					
25.5					
27.5					
29.5					

Table 4.17

If you have access to a spreadsheet, produce your own completed version of this table. Otherwise copy Table 4.17 and complete it using your graphics calculator.

4 Write a null hypothesis and an alternative hypothesis.

5 Use these expected frequencies in a χ^2 test of the suitability of a normal distribution.

Experiment with merging different class intervals. What happens to the values of $\sum\frac{(O-E)^2}{E}$ and ν if you merge four cells at each end? What is the conclusion from a χ^2 test if you do this?

What is different if you merge three cells at each end? Is the conclusion different in this case? Think about how large a value of X^2 is required for the result to be significant in each of these cases.

6 Decide which is the best way to carry out the χ^2 test in this case, and write a short paragraph explaining your conclusions.

Normal probability graph paper

Large data sets

Probability graph paper is specially designed to test whether data could come from a normal distribution. If you plot a cumulative frequency curve for a normal distribution, you should obtain a smooth S-shaped curve. However, it can be difficult to spot deviations from this curve when a skewed distribution is plotted.

To make the process easier, special probability graph paper has been designed which stretches out the cumulative frequency graph for a normal distribution so that it becomes a straight line (Figure 4.3).

Activity 4.12 Using normal probability graph paper

1 Draw up a cumulative frequency table for the reaction times data from Activity 4.11.

2 Convert the cumulative frequencies to cumulative relative frequencies which you can treat as cumulative probabilities. Many probability graph papers have the vertical cumulative probability axis expressed as a percentage rather than as a decimal.

3 To draw your own normal probability graph, first make your own scale across the horizontal axis to accommodate your data measurements. Make a vertical scale for the cumulative probabilities.

On the horizontal axis, plot the upper limits for each class interval, and on the vertical axis plot the cumulative relative frequency.

4 The points lie fairly close to a straight line. By eye, draw a straight line through the points.

5 From your graph, estimate the mean of the distribution, and the standard deviation. Explain why you can use the median as a convenient estimate of the mean. A good way to estimate the median is to draw a line across the graph at 50% and read the median value from the horizontal axis. Compare this estimate with your answer to question **2** in Activity 4.11.



6 Explain how you can use the 84th percentile to give you an estimate of the standard deviation, and find such an estimate.

Activity 2.18 Euthanasia

1 Table 4. 18 is a frequency table which gives the pulse rates of a group of female students.

Beats per minute	Number of students
50 – 53	1
54 – 57	1
58 – 61	5
62 – 65	12
66 – 69	23
70 – 73	25
74 – 77	20
78 – 81	13
82 – 85	4
86 – 89	1

Table 4.18

H Strictly speaking, Table 4.18 shows discrete data but you will have to treat the data as continuous.

Use normal probability graph paper to test whether the normal distribution is a suitable model for this set of data. Estimate the mean and standard deviation from your graph.

2 How will non-normal data appear when plotted on normal probability graph paper? If you have access to a statistical package on a computer, use it to generate some data from other distributions, and see how these look when you plot them on normal probability graph paper.

Making your own normal probability graph paper is quite easy. Activity 4.14 will help you to understand how the paper works.

Activity 4.14 Making probability graph paper

1 Take a piece of ordinary graph paper and mark the vertical axis from –3 to +3. Use a pencil for this as you will need to rub it out later. This scale represents Z.

2 Now refer to normal distribution tables (Table 1, Appendix B) to find the area to the left for each value of Z on your scale. Mark these in ink on your vertical axis, removing the Z values as you go. Figure 4.2 shows how to do this.

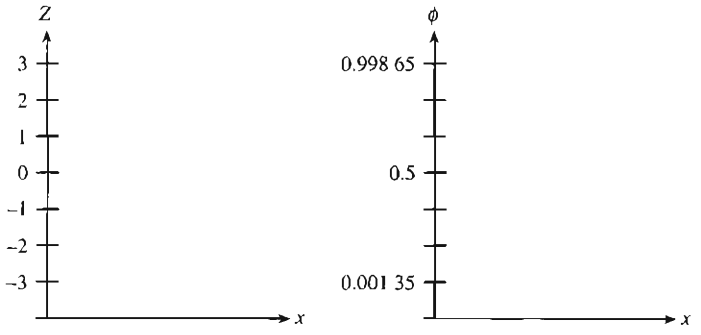


Figure 4.2

In the diagrams $\phi = 0.5$ corresponds to $Z = 0$, and $\phi = 0.99865$ corresponds to $Z = 3$.

Small data sets

For small data sets, plotting the results on normal probability graph paper is the only way to test the goodness of fit to a normal distribution. However, because there are only a few data points, some adjustment has to be made. Here are some results obtained from the friction experiment, described in Activity 14.1 of the *Distributions* unit in Book 5. They are arranged in order:

54, 54, 56, 57, 62, 63, 64, 64, 66, 68, 69, 69, 71, 73, 76.

These measurements represent the height in mm of the plank at the point where the wooden block begins to slide down. While these are measurements made on a continuous ratio scale, there are so few of them that they appear as discrete points. With 15 points, the object is to divide up the range $-\infty$ to $+\infty$ into 16 intervals, using the results as the boundaries of these intervals.

Method of working

1 Count the number of results ($= n$).
2 Then work with $n+1$ intervals.
3 The cumulative relative frequency for each value of x_i for $i = 1, 2, 3, \ldots, n$ is $\frac{i}{n+1}$.

x	Cumulative relative frequency $\frac{i}{n+1}$		
54	$\frac{2}{16}$	0.125	12.50%
54			
56	$\frac{3}{16}$	0.1875	18.75%
57	$\frac{4}{16}$	0.25	25.00%
62	$\frac{5}{16}$	0.3125	31.25%

Table 4.19 *(Continues overleaf)*

63	$\frac{6}{16}$	0.375	37.50%
64	$\frac{8}{16}$	0.5	50.00%
64			
66	$\frac{9}{16}$	0.5625	56.25%
68	$\frac{10}{16}$	0.625	62.50%
69	$\frac{12}{16}$	0.75	75.00%
69			
71	$\frac{13}{16}$	0.8125	81.25%
73	$\frac{14}{16}$	0.875	87.50%
76	$\frac{15}{16}$	0.9375	93.75%

Table 4.19

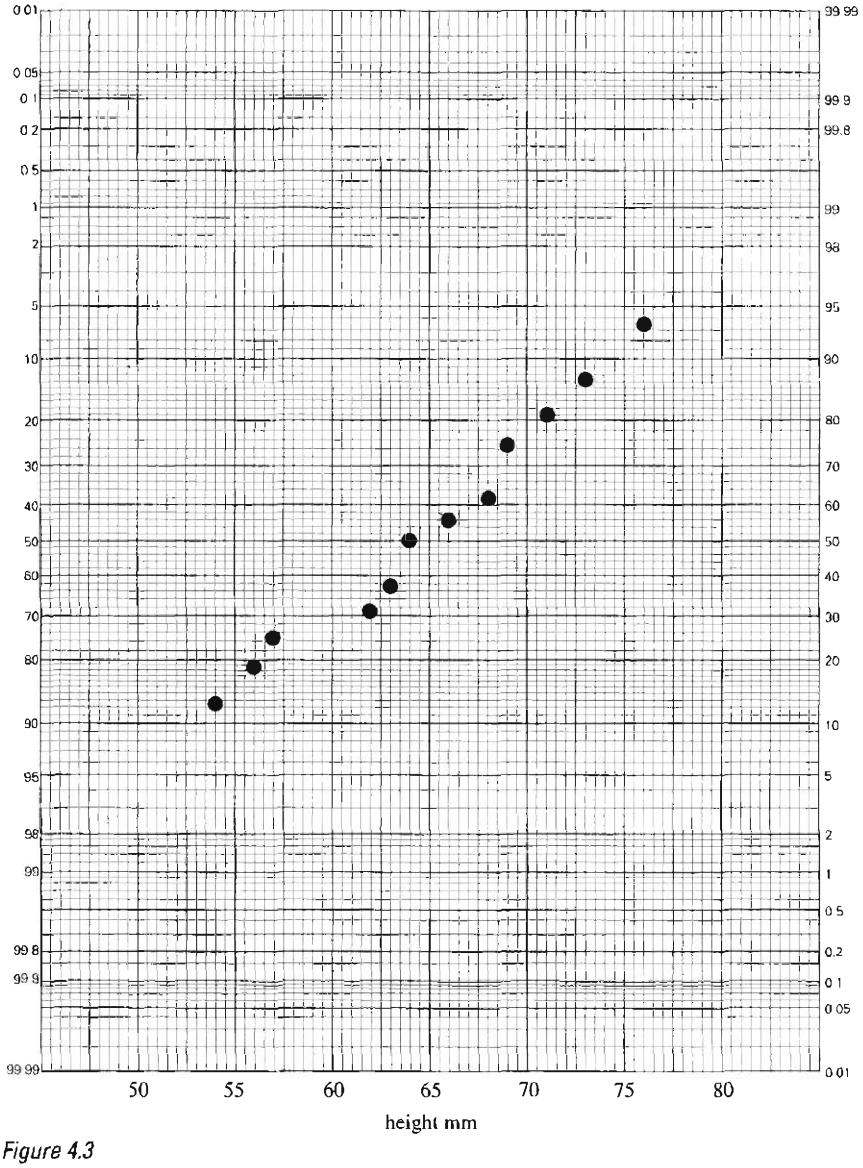


Figure 4.3

The graph has been drawn, showing that the first three points seem to be 'out of line' from the others. A line can be drawn through the other points fairly well.

Activity 4.15 Estimating mean and standard deviation

1 Estimate the mean and standard deviation from the graph.

2 Use your calculator to find the mean and standard deviation.

Checklist for deciding whether the normal distribution is a suitable model

Large data sets only $(n \geq 30)$

1 Draw a histogram.
Is the distribution uni-modal?
Is the distribution symmetrical?
2 Calculate the mean and standard deviations.
Is the mean close to the mode?
3 Is the variable continuous, or is it reasonable to assume that the variable is continuous?
Are the measurements on an interval scale?
4 For a χ^2 test:
ν = number of classes − 3.
5 Use normal probability paper.

Small data sets $(n < 30)$.

A histogram is unlikely to help much. Plotting the data on normal probability paper is the only possible test.

Plot points at $\left(x_i, \frac{i}{n+1}\right)$.

Do the points lie approximately on a straight line?

Using χ^2 to test association

In the *Correlation and prediction* unit in Book 3, you found some techniques for measuring the correlation between two variables x and y. All these methods require x and y to be at least ordinal or ratio quantities. In this section, you will find ways of testing for association between two categorical variables using the chi-squared test.

Activity 4.16 Chi-squared test of association

A student has collected some data to see whether vegetarian students are less likely than non-vegetarians to smoke. She thinks that vegetarians may be more health conscious than other students. Table 4.20 shows her results set out in a two-way table (a **contingency table**).

	Vegetarian	Non-vegetarian
Smokers	5	21
Non-smokers	27	42

Table 4.20

1 Write down a null hypothesis and an alternative hypothesis for her study.

2 If the null hypothesis is true, you would expect the smokers to be distributed equally among the vegetarians and the non-vegetarians. Unfortunately, there are not equal numbers of these two groups. Table 4.21 shows the survey results (to be called **observed frequencies** from now on) with some of the totals. These totals are all called **marginal totals**.

Observed frequencies	Vegetarian	Non-vegetarian	Totals
Smokers	5	21	
Non-smokers	27	42	
Totals	32		95

Table 4.21

Copy and complete the table.

3 In the same way that you used χ^2 to test for significant differences between observed frequencies and expected frequencies in the first half of this chapter, you can use a similar method here if the expected frequencies can be found.

You must derive the expected frequencies from the null hypothesis in some way, so that the numbers of smokers are distributed in the right proportions among vegetarians and non-vegetarians. Also, the expected frequencies must give the same marginal totals as the observed frequencies.

The way to do this is to draw up a table of expected frequencies, working to one decimal place. Table 4.22 shows how to start. Fill in the rest of the marginal totals from your answers to question 2.

Expected frequencies	Vegetarian	Non-vegetarian	Totals
Smokers			
Non-smokers			
Totals	32		95

Table 4.22

The best estimate of the proportion of all people who are vegetarian is $\frac{32}{95}$. In the same way, you can estimate the proportion of the population who are non-vegetarians.

Use these proportions to estimate the numbers of smokers in the sample who would be expected to be vegetarian and non-vegetarian if the null hypothesis is true. Do the same for the non-smokers.

4 Now that you have the expected frequencies, you can use χ^2 to test for goodness of fit between observed frequencies and expected frequencies using the same formula as before.

$$X^2 = \sum \frac{(O-E)^2}{E}$$

It is helpful to set out your working in a table such as Table 4.23.

Observed frequency	Expected frequency	(O – E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	8.8	–3.8	14.44	1.64
27				
21				
42				

Table 4.23

5 In order to test whether this value of $\sum \frac{(O-E)^2}{E}$ is unusually high within the χ^2 distribution, you need to know which χ^2 distribution is to be used. Remember that you can think of the number of degrees of freedom as the number of frequencies that can be chosen freely. Look at the two-way table in question **3** with the marginal totals fixed. Decide how many of the frequencies can be freely chosen, and how many are determined by the marginal totals.

6 **a** Refer to tables to find the significant value of χ_1^2 (chi-square with one degree of freedom).
b Is the result of this survey significant?
c What are your conclusions?

7 **a** Try a two-way contingency table with three rows and two columns. The dots in Table 4.24 represent known totals.

	C1	C2	Totals
R1			•
R2			•
R3			•
Totals	•	•	•

Table 4.24

How many degrees of freedom are there?

b Now suppose that the table has three rows and three columns. How many degrees of freedom are there?

What if there are four rows and two columns? What about four rows and three columns? Make a generalisation.

How would you justify your generalisation?

Activity 4.17 More practice with the chi-squared test

For each problem write down:

a the null hypothesis

b the alternative hypothesis

c the frequencies expected under the null hypothesis

d the value of $\sum \frac{(O-E)^2}{E}$

e the number of degrees of freedom

f whether your result is significant

g your conclusions.

1 A student conducts a survey to see whether people tend to wear their watch on the hand that they do not use for writing, or whether they will wear it on either hand. Table 4.25 shows her results.

	Watch on right hand	Watch on left hand
Writes with the right hand	11	39
Writes with the left hand	12	8

Table 4.25

2 To determine the usefulness of a new measles vaccine, an experiment is set up with 200 children. The children are allocated at random to two groups of 100. The children in one group receive the vaccine but the children in the other, control, group receive only the inert diluting agent, in which the vaccine is normally administered. After one year the numbers of children who have contracted measles in both groups are counted, and the results are shown in Table 4.26.

	Caught measles	Have not caught measles
Vaccinated group	10	90
Control group	34	66

Table 4.26

Is the new vaccine effective?

3 A survey of hospital staff is conducted to determine levels of job satisfaction in different categories of staff. The findings are summarised in Table 4.27.

	Satisfied	Dissatisfied
Doctors	73	7
Nurses	45	45
Auxiliaries (unqualified staff)	22	68

Table 4.27

4 A scientist studying various inherited characteristics examines the relationship between hair colour and eye colour. The results are shown in Table 4.28.

Hair colour	Eye colour		
	Blue	Brown	Black
Black	12	21	9
Brown	21	54	13
Red	5	3	3
Blonde	19	8	4

Table 4.28

Since several expected frequencies are below 5, what do you think should be done: eliminate those with red hair or combine them with another hair colour?

Are hair colour and eye colour related?

You could survey students from your own school to see if their hair and eye colours are consistent with this statement.

Activity 4.18 Observation of body language

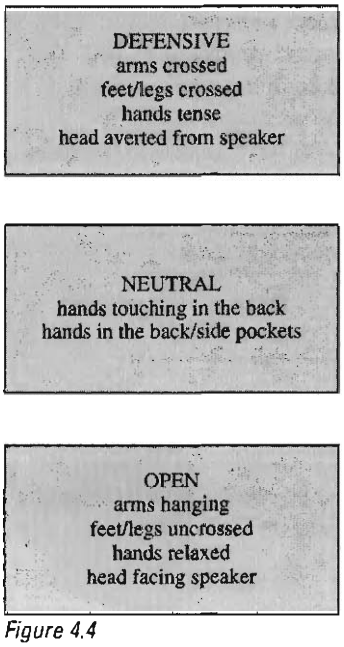


Figure 4.4
Categorising postures

This is an observation study of body language, providing data which you can analyse using a χ^2 test.

Non-verbal communication has been found to be especially important in people's social interaction. We convey many of our feelings through facial expressions, gestures and body movements. Certain postures are taken to be defensive, expressing a feeling of vulnerability and a desire to defend physical space. If people feel threatened, they tend to fold their arms, thereby protecting their body, or cross their legs. A person who does not feel threatened will tend to adopt an open posture.

Here are various hypotheses that you can investigate. Choose one of them.

a Are females more likely to adopt defensive postures?

b Are younger people more likely to adopt defensive postures?

c Is body language related to status? Can you observe interactions between teachers and students, say? You will need to be discreet with this one!

Whichever hypothesis you choose, you must collect observations for at least ten people in each of two groups:

- females and males
- younger and older people

- teachers and students
- high status and low status.

In order to collect your observations, you need to find a place where you can be unobtrusive, and to work in such a way that people are not aware that they are being watched.

As people will possibly behave differently when they are sitting or standing, collect observations for subjects who are all sitting, or all standing. If you have time you can do two surveys, one for 20 subjects sitting and one for 20 subjects standing.

You will need a carefully-designed observation sheet to help you to record the data with the minimum amount of fuss, and to help you when you come to analyse the results later. A suggested observation sheet is given below, but you may like to improve on it.

To analyse your data, count up the ticks on your observation sheet, for open posture and closed posture for each group of subjects. Make a contingency table with the format of Table 4.29 (neutral postures being ignored).

	Females	Males
Defensive		
Open		

Table 4.29

Use a χ^2 test to test your null hypothesis.

Subjects are.....*

	Subject number									
	1	2	3	4	5	6	7	8	9	10
ARMS										
hanging										
crossed										
neutral										
LEGS										
feet/legs uncrossed										
feet/legs crossed										
HANDS										
relaxed										
tense										
HEAD										
averted from speaker										
facing speaker										

Tick each category observed.
* Fill in female/male, low status/high status

Figure 4.5
Suggested observation sheet for observing body language

Summary – a test of association

- A χ^2 test can be used to test for an association between two categorical variables.
- A two-way table, or contingency table, must show frequencies, not proportions or measurements.
- Expected frequencies can be calculated by:

$$\text{expected frequency} = \frac{\text{column total} \times \text{row total}}{\text{grand total}}$$

- Expected frequencies much below five are best avoided, but some discretion may be used.
- The number of degrees of freedom for an $m \times n$ contingency table is given by $\nu = (m-1)(n-1)$.
- Cautionary note: a significant value for χ^2 shows that there is an association between the two variables. It does not necessarily show that changes in one variable cause changes in the other.

Reflecting on Chapter 4

What you should know

- how to use the χ^2 distribution to test the fit of a given model
- how to fit a binomial model to a frequency distribution, and how to test the fit of your model
- how to fit a normal model to a frequency distribution, and how to test the fit of your model
- how to use normal probability graph paper to explore the possibility of fitting a normal model to a set of data
- how to use the χ^2 distribution to test the association between two nominal variables.

Preparing for your next review

- Think about how you might use the ideas you have studied in this chapter to analyse the data from your own investigation.
- Answer the following check questions.

1 Look at the data generated by student G in Table 4.1. Fit a binomial distribution to student G's results and test the fit of this model.

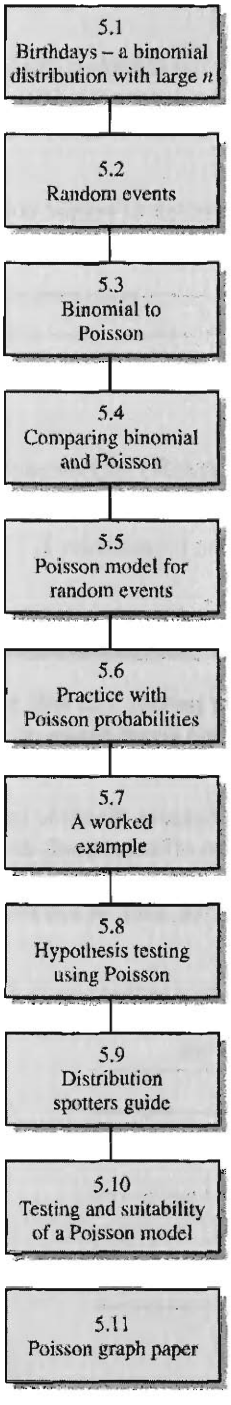
2 Table 4.30 shows the results of a study of insects collected on different backgrounds and grouped according to colour. Test whether the colour of an insect is associated with the background on which it is found.

Colour of insect	Background	
	Grass	Flowers
Black/Grey	29	47
Coloured	39	30

Table 4.30

Practice exercises for this chapter are on page 102.

5 The Poisson distribution



In this chapter you will learn about the Poisson probability distribution, which you can use to model situations of a particular type. Examples include: the number of telephone calls received in a telephone exchange between 9 a.m. and 10 a.m. every weekday; the number of arrivals per minute in a post office; the number of goals in a football match; the number of flaws in a metre length of cloth; the numbers of a particular kind of insect which can be found in a square metre of grass in a field.

Work through the activities in sequence.

In Activity 5.1, you will find it helpful to share the work and combine your results with other statistics students.

Activity 5.3 requires the use of a spreadsheet.

Activity 5.11 is an optional activity, involving the use of Poisson graph paper.

The Poisson distribution is related to the binomial distribution, and this relationship is explored through the data collected in Activity 5.1.

Activity 5.1 Birthdays – a binomial distribution with large n

If possible, work with another statistics student, so that you can share out the work in this activity.

In a sixth form college of 300 students, is it likely that another person's birthday occurs on the same day as yours? What is the chance that five other people in the college share your birthday?

1 Use the class registers for your school or college, or the school or college student database if you are allowed access to some of it, to find out, for each day of the calendar year, how many of the students in a sample of 300 have their birthday on that date.

2 From your results, make a frequency table for the number of people born on the same day. Set it out as in Table 5.1.

Number of people born on the same day	0	1	2	3	4
Frequency					

Table 5.1

3 What is the greatest number of people born on any one day in your sample of 300? How often does that number occur?

4 Draw a bar-line graph of the results that you obtained for question **2**.

Activity 5.2 Random events

In this activity you need to work with at least one other person. You will need a digital watch or stop-watch, a clipboard, and a recording sheet drawn up in advance.

Go with your partner to a place where people arrive at random. Possible locations are: a petrol station, a pedestrian crossing, a bank, a post office, a small shop or café, or the entrance to a car park. When you get there, you need to count the number of people (or cars) arriving in a fixed time interval, such as one minute or half a minute.

1 Before you go, make a data collection sheet for 60 time periods, as in Table 5.2.

Time period	Number of arrivals
1	
2	
3	
⋮	
60	

Table 5.2

One person must time intervals of 1 minute. The other person counts the number of arrivals at the observation point. Collect observations for 60 equal time intervals. You will find that you get the best data at a fairly quiet time of the day. If you collect data at a very busy place, it is better to work in shorter time periods – 30 seconds or even 15 seconds – so that your maximum count in any time period is well below ten.



2 When you get back, make a frequency table of your results.

3 Draw a bar-line graph for your results.

4 Do you have discrete data or continuous data?

The Poisson distribution as a model for random events

Look at the frequency table for your results for Activity 5.2, and consider what distribution you might use as a model in this situation. Is the binomial model suitable?

Although the results are for a discrete variable, and the number of arrivals in each time interval is recorded, you cannot say how many 'non-arrivals' there are. You cannot give a value either to the binomial parameter n, since for any time interval there is no maximum possible number of arrivals.

However, you can calculate the mean from this set of data. So although you cannot say what n is, you can estimate np from the mean number of arrivals per minute.

Imagine a one-minute time interval divided into n small time intervals, so that there cannot be more than one arrival in any small interval. It is now possible to give meaning to the probability p of an arrival in any small interval:

$$\text{P(an arrival in one small interval)} = \frac{\text{mean number of arrivals per minute}}{\text{number of small intervals in one minute}} = \frac{\mu}{n}$$

so $p = \frac{\mu}{n}$ and $q = 1 - \frac{\mu}{n}$. In this case, the number μ is fixed.

Using a binomial model, the probability of x arrivals per minute, $\mathrm{P}(X = x)$, is given by

$$\mathrm{P}(X = x) = \binom{n}{x}\left(\frac{\mu}{n}\right)^x\left(1 - \frac{\mu}{n}\right)^{(n-x)}.$$

But this expression is awkward to use. How does it behave as n becomes large and p becomes small, in such a way that the product np is fixed and $np = \mu$?

The answer is that $\binom{n}{x}\left(\frac{\mu}{n}\right)^x\left(1 - \frac{\mu}{n}\right)^{(n-x)}$ behaves like $\frac{\mathrm{e}^{-\mu}\mu^x}{x!}$ when n is large, but it is not easy to prove this approximation theoretically. However, in Activity 5.3 you can see how large n needs to be for the result to be convincing.

Activity 5.3 Binomial to Poisson

You will need a spreadsheet for this activity.

1 Suppose that $\mu = 5$. Use your spreadsheet to calculate the values of $\binom{n}{x}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{(n-x)}$ and $\frac{e^{-\mu}\mu^x}{x!}$ for $n = 10$ and $p = 0.5$. Draw graphs of these expressions for values of x from 0 to n inclusive. You should see some differences between your graphs, although they will have the same general shapes.

2 Now increase the value of n to 20, and decrease p so that $np = 5$. Draw the graphs again, and compare them.

3 Continue increasing the value of n by 10 until you can no longer distinguish between the graphs of $\binom{n}{x}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{(n-x)}$ and $\frac{e^{-\mu}\mu^x}{x!}$. Note down the values of n and p when this happens. Compare your values of n and p with those obtained by other students.

You might expect this new distribution to be similar to the binomial distribution whenever n is large and p is small. In Activity 5.4 you will check this against the data from Activity 5.1.

Activity 5.4 Comparing binomial and Poisson

1 Calculate probabilities for the birthdays example in Activity 5.1, using both binomial and Poisson models as far as $x = 4$.

x	Probabilities for binomial model $p = \frac{1}{365}$	Probabilities for Poisson model
0	$(1-p)^n$	$e^{-\mu}$
1	$np(1-p)^{n-1}$	$\mu e^{-\mu}$
2	$\frac{n(n-1)}{2!}p^2(1-p)^{n-2}$	$\frac{\mu^2}{2!}e^{-\mu}$
3	$\frac{n(n-1)(n-2)}{3!}p^3(1-p)^{n-3}$	$\frac{\mu^3}{3!}e^{-\mu}$
4	$\frac{n(n-1)(n-2)(n-3)}{4!}p^4(1-p)^{n-4}$	$\frac{\mu^4}{4!}e^{-\mu}$

Table 5.3

Notice that, for this binomial model, n is the number of students in your survey and $p = \frac{1}{365}$. Hence, for this Poisson model, $\mu = \frac{n}{365}$.

2 Compare the values of your binomial and Poisson probabilities. How close are they?

3 Compare both sets of probabilities to your survey results. In order to make this easier, multiply your probabilities by your total frequency (n in this case) to give theoretical expected frequencies.



4 How close are your results to the theoretical expected frequencies? Use the χ^2 test that you studied in Chapter 4 to test the fit of each of the two models that you have here. Remember that you will need to group together the values of x so that no interval has a very small theoretical expected frequency.

5 Read again the assumptions behind the binomial model. Write brief reasons explaining why they might not be valid assumptions for the birthdays data.

Activity 5.5 Poisson model for random events

1 Calculate the mean of your frequency table for your results of Activity 5.2.

2 Use your value for the mean (μ) to calculate Poisson probabilities for $x = 0, 1, 2, 3, 4, 5, 6, \ldots$ arrivals

$$P(X = x) = \frac{\mu^x}{x!} e^{-\mu}.$$



3 Calculate the theoretical expected frequencies for the Poisson model, and compare them with your observed results.

4 Draw a bar-line graph for the expected frequencies. Compare this with the bar-line graph that you drew for Activity 5.2.



5 Use the χ^2 test to test the fit of the Poisson distribution as a model for your results.



6 If your results are not close to the theoretically expected frequencies, consider whether any of the assumptions or conditions for the Poisson distribution, given below, have been broken.

Conditions for modelling with a Poisson distribution

- Events (such as arrivals) occur at random in a continuous space or time period.
- Events occur singly; it is impossible for two events to happen simultaneously.
- Events occur uniformly, so that the expected number of arrivals is proportional to the size of the time interval. The mean number of arrivals in two minutes, for example, should be twice the mean for one minute. This mean should be constant from day to day at that time.
- Events occur independently; the probability of an arrival in any small interval is independent of arrivals in other intervals.
- The variable is discrete; it is a count of the number of events that occur in a sample space of a given size.

Examples of possible sampling spaces in different contexts include

- a time interval
- an area of cloth, or
- a volume of liquid.

Examples of situations in which a Poisson distribution may be used as a model are

- the number of phone calls arriving at a switchboard in a given time interval
- the number of white blood cells in a given volume of blood
- the number of small organisms in a given volume of water
- the number of particles emitted by a radioactive source in a small time interval. Generally background radiation is sufficiently high that a radioactive source is not required, and clicks on a Geiger counter should give a Poisson distribution
- the number of flaws in given lengths of cloth.

Vehicles passing an observation point in a town centre are unlikely to be modelled by a Poisson distribution. This is because traffic lights, roundabouts and other obstructions will cause traffic to bunch rather than flowing freely.

Example 1

On past experience, the staff in a small car insurance company believes that the average number of accident claims a year should be five. Assuming that this is correct, use the Poisson distribution to calculate the probability of having fewer than three claims in a year.

Solution

The average number of claims per year is five. Assume that the number of claims per year follows a Poisson distribution with mean $\mu = 5$.

You want the probability of fewer than three claims, which is

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-5} + e^{-5} \times 5 + e^{-5} \times \frac{5^2}{2} \\ &= 0.0067 + 0.0337 + 0.0842 \\ &\approx 0.125 \text{ to 3 decimal places.} \end{aligned}$$

Notation

The Poisson distribution with mean μ is often written $\text{Po}(\mu)$. So, in Example 1, the distribution would be written as $\text{Po}(5)$.

Example 2

In the car insurance company described in Example 1, calculate the probability of getting six or more claims in six months.

Solution

The average number of claims in six months is 2.5. So the distribution of the number of claims in six months is Po(2.5). You want to find

$$
\begin{aligned}
P(X \geq 6) &= 1 - P(X \leq 5) \\
&= 1 - \left(P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)\right) \\
&= 1 - \left(e^{-2.5} + e^{-2.5} \times 2.5 + e^{-2.5} \times \frac{2.5^2}{2!} + e^{-2.5} \times \frac{2.5^3}{3!} + e^{-2.5} \times \frac{2.5^4}{4!} + e^{-2.5} \times \frac{2.5^5}{5!}\right) \\
&= 1 - (0.9580) \\
&\approx 0.042 \text{ to 3 decimal places.}
\end{aligned}
$$

Activity 5.6 Practice with Poisson probabilities

1 The mean number of telephone calls received in ten-minute intervals at a college switchboard is five. Assuming that the Poisson model applies in this situation, calculate the probability that in ten minutes

a no calls are received

b only one call is received

c four or more calls arrive.

2 In two hours, 3600 particles are emitted by a radioactive material. Find the mean for a ten-second interval. Use this value to calculate Poisson probabilities for $(0, 1, 2, \ldots, 12)$ particles emitted per ten-second interval.

3 A company finds that on average two faults occur in its computer system every month. Find the probabilities of

a four

b five or more

faults occurring in a month.

What assumptions have you had to make?

4 A college finds that, on average, three of its students gain industrial scholarships each year. Find the probability of the college winning

a three scholarships

b more than three scholarships

in a year.

Do you think the Poisson model is appropriate here?



5 The library caretaker finds that, on average, four light bulbs have to be replaced every week. She replenishes the stock of light bulbs weekly; how large should this stock be to ensure that the probability of not having enough is less than 0.01?

The Poisson distribution as a model for hypothesis testing

Activity 5.7 A worked example

In a chemical plant, the average number of industrial accidents per month is three. New working practices are introduced following recommendations made by the

health and safety officer. In the following month there is only one accident. Is this a significant decrease?

1 First start by writing down a null and an alternative hypothesis.

2 Now use the value of the mean in your null hypothesis to calculate the required probabilities. Draw a diagram to show these probabilities.

3 Are these results unusual?

Activity 5.8 Hypothesis testing using Poisson

For each of the following problems, write down:

- the null hypothesis H_0
- the alternative hypothesis H_1
- the relevant probabilities
- your conclusion.



1 At a particular mini roundabout, there are 12 accidents on average per year. Eventually this roundabout is changed back to a junction with traffic lights. In the next six-month period there are just two accidents. Is this a significant reduction?



2 On average, the local gas board receives one phone call each day reporting a gas leak. One night there is a sharp frost, the following morning there are four calls. Is this a significant increase?

3 A restaurant owner usually finds two bad tomatoes in the daily consignment that she receives from her supplier. She has to change her supplier, and the following day finds six bad tomatoes. Is this a significant increase?



4 A student receives an average of two letters a week. One day he gets five letters. Is this really unusual?

Activity 5.9 Distribution spotters' guide

Use a spreadsheet or a graphics calculator to work out the probabilities in the table below.



1 It is best to use an iterative method to calculate the Poisson probabilities when you need a sequence as you do here. This means using the value that you have calculated for $P(X = x - 1)$ to calculate the value of $P(X = x)$. Find an equation for $P(X = x)$ in terms of μ, x and $P(X = x - 1)$.

2 Copy and complete Table 5.4 by calculating the appropriate Poisson probabilities for different values of μ.

x	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$
0					
1					
2					
⋮					
10					

Table 5.4

3 Also calculate probabilities for $\mu = 10$ for appropriate values of x.

4 For each value of μ, draw a graph to show the probabilities.

Activity 5.10 Testing the suitability of a Poisson model

A group of students organises a survey to count the number of customers arriving each minute at a small post office. Table 5.5 shows the results:

Number of customers arriving per minute	0	1	2	3	4	5
Frequency	25	20	13	0	1	1

Table 5.5

1 What conditions are needed for a Poisson distribution to apply in a particular case?

2 What are the parameters needed to calculate probabilities using a Poisson model?

3 Draw a bar-line graph to illustrate the data. Describe the shape of the distribution.

4 Calculate the mean and variance. Are these similar in value?

5 Use the Poisson distribution to generate theoretical probabilities as shown in Table 5.6.

The probabilities are given by the formula: $\mathrm{P}(X = x) = \frac{\mu^x}{x!}\mathrm{e}^{-\mu}$.

x	0	1	2	etc.
$\mathrm{P}(X = x)$	$\mathrm{e}^{-0.92}$	$0.92\mathrm{e}^{-0.92}$	$\frac{0.92^2}{2}\mathrm{e}^{-0.92}$	etc.

Table 5.6

6 Multiply each probability by the total frequency, which is 60, to find the expected frequencies.

7 Write a null hypothesis and an alternative hypothesis.

8 In this situation there are two constraints on the expected frequencies. What are they?

9 It is suggested that you use the following classes for calculating χ^2: $x = 0$, $x = 1$, $x = 2$ and $x \geq 3$, even though the group '$x \geq 3$' has an expected frequency of only 4. If the table was reduced down further to only three classes there would be only one degree of freedom, and the frequency distribution would be obscured.

10 Use tables to find the 5% cut-off value for χ^2.

11 Decide whether your result is significant, and carefully state your conclusions.

Activity 5.11 Poisson graph paper

This activity is optional.

Table 5.7 gives the results used in Activity 5.10, showing the number of customers arriving at a small post office in each of 60 one-minute intervals.

Number of customers arriving per minute (x)	Frequency
0	25
1	20
2	13
3	0
4	1
5	1

Table 5.7

In order that the data can be plotted, they must be re-organised as in Table 5.8.

Cumulative relative frequencies calculated from the 'bottom upwards'

x	Probability of observing this value of x or greater
0	$\frac{60}{60} = 1$
1	$\frac{35}{60} = 0.583$
2	$\frac{15}{60} = 0.250$
3	$\frac{2}{60} = 0.033$
4	$\frac{2}{60} = 0.033$
5	$\frac{1}{60} = 0.017$

Table 5.8

For each value of x, you need the probability of observing that value of x or greater. You can find this probability by finding cumulative frequencies working upwards from the bottom of the table, and converting these to relative frequencies.

Poisson graph paper

At first glance this looks very complicated, as there are three sets of lines. The cumulative probabilities are shown on the vertical scale:

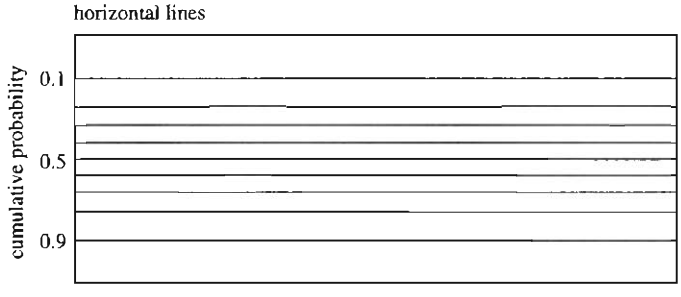


Figure 5.1

Values of x are plotted on the curved lines:

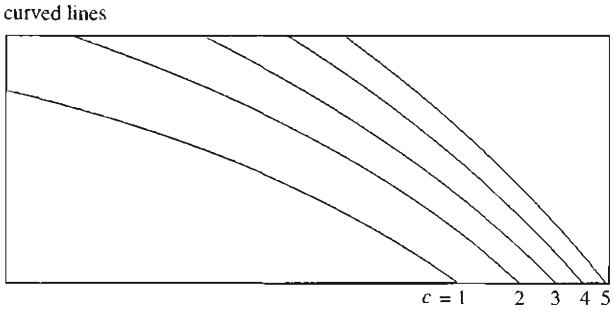


Figure 5.2

Ignore the vertical lines when you are plotting points; these are used later. Here is the post office data plotted on the Poisson graph paper.

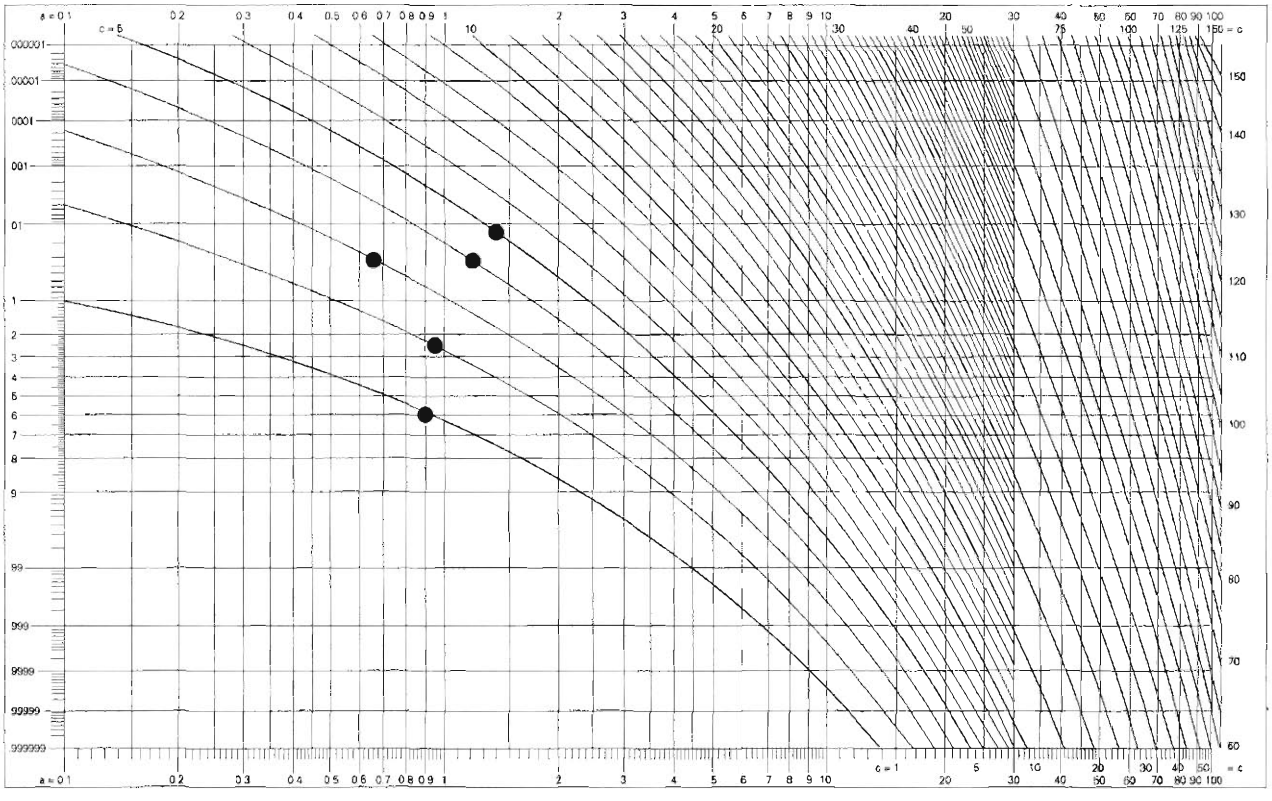


Figure 5.3

If the Poisson distribution is a suitable model, the points should lie close to a vertical straight line.

For the post office data, the values for $x = 1$, 2 and 3 do lie more-or-less on a straight line. The points for $x = 4$ and 5 do not seem to fit the pattern predicted by the model. But these represent only two observations out of 60. If the points have fallen on a straight vertical line the value of the mean can be read off on the horizontal axis, using those straight vertical lines which have been ignored so far.

Notice, though, that the points for $x = 1$, 2 and 3 do lie close to the line indicating a mean at around 0.9.

The number of phone calls received in a small office, in each of 17 consecutive half-hour periods:

2, 2, 1, 3, 2, 3, 4, 3, 2, 0, 3, 4, 1, 3, 2, 3, 2.

1 Form a frequency table for the results given in this question, and perform suitable calculations to plot the data on Poisson graph paper.

2 Estimate the mean from the graph.

3 Compare this with the value of the sample mean obtained by direct calculation.

Checklist for a Poisson distribution

1 Do you have discrete data?
2 Draw a graph. Is the distribution positively skewed?
3 Are the events random? Are they independent?
4 Calculate the mean and the variance. Are they roughly equal in value?
5 For a χ^2 test: ν = number of classes − 2.

Reflecting on Chapter 5

What you should know

- how to calculate probabilities using a Poisson distribution
- how to fit a Poisson model to a frequency distribution
- how to use Poisson probability paper to explore the possibility of fitting a Poisson distribution to a set of data.

Preparing for your next review

- Answer the following check questions.

1 For each of the following situations, decide whether or not a Poisson model is appropriate, explaining your decision and listing any assumptions that you have made. If you decide that a Poisson distribution is not appropriate, consider whether a normal or a binomial model is appropriate. If you can justify a suitable model, solve the problem.

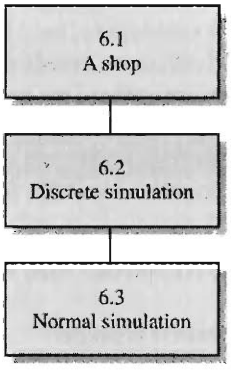
a Cars arrive at a petrol station at an average rate of two per minute. What is the probability of six cars arriving in a two-minute period?

b Imperfectly printed fabric is on sale at a local market at a special price. The market trader honestly admits that there is, on average, one printing error on every metre of fabric. What is the probability that there will be four errors or more on a length of six metres?

c A student has analysed the results of a local football league for 12 months, and has found that home teams have scored a total of 265 goals in 90 matches. Next weekend there will be nine matches. What is the probability that all the home teams will score at least one goal each?

Practice exercises for this chapter are on page 103.

6 Simulations



Work through the activities in order.

With the rise in the power and accessibility of computers, simulations have become increasingly important statistical methods. This is especially true of stochastic simulations, which make use of the computer's pseudo-random number generator to simulate random processes. You were introduced to this idea in the *Modelling with probability* unit in Book 2. Here are two examples of the use of stochastic simulation in research, each using some of the ideas that you have learned about distributions in this option and in other units.

An operational research manager is asked to estimate how long it will take to complete a multi-million pound project. The manager does not have the expertise in all aspects of the project to be able to answer this question, so decides to consult the project contractors responsible for each section of the project. The contractors are asked to use their experience of similar projects to give three estimates of the time required: a shortest possible time if everything goes smoothly; a most likely time; and a longest time. These three estimates are used to construct a probability distribution based on the normal distribution for the time required for each section of the project. The probability distributions are then used in a 'Monte Carlo' simulation of the project to estimate the likely time required, and the risk of the project taking too long.

Large sums of money are at stake in projects of this kind, and, because of the inter-relationship between different tasks, the scheduling of the work to be done by the different contractors is complicated. A Monte Carlo simulation of this kind is sometimes the only effective way to take account of all the available information and experience to arrive at an overall picture. The method has the advantage of providing not only a basis for estimating the most likely time required, but also a way to work out the financial risks involved.

In another example, a team of university researchers conducted a study of the use of a pedestrian crossing, and collected a large amount of data. After analysing the data, they used the results to construct a model of the interaction between road traffic and the pedestrians using the crossing. This model included the use of a Poisson distribution for the number of cars arriving at the crossing in a minute. The model was used to build a stochastic simulation, and the data generated by the simulation was used to modify the model. So simulation can be thought of as a source of data, alongside experiments and surveys.

You may have already generated data using simulations, in earlier units of the course as well as in activities suggested in this option. Now you may wish to embark on a longer project and design a simulation yourself.

Here are further examples of the kinds of questions you can answer by using a simulation.

- What is the ideal number of check-outs in a supermarket?
- Is a central queue a better system than queues at separate counters in places such as banks or post offices?
- How many potential customers walk away because the queue extends out of the doorway of a shop onto the pavement?
- What is the optimum number of assistants in a baker's shop?

Animal and bird populations are interesting subjects for simulations, but you need access to reliable and relevant data such as birth rates, death rates, breeding patterns. If you can find data already collected from an expert source, rather than having to gather your own observations, it may be possible to simulate the entire population of a threatened species or a subset of the population. Ornithologists are very enthusiastic data gatherers, and there is a wealth of information available for simulations of bird populations.

A simulation is a model of a real situation. You can use it to predict what will happen if changes are made. For example

- will people wait for shorter times if the queuing system is changed?
- will the manager increase his profits by employing another shop assistant?
- how much more quickly will a disease spread if fewer children are inoculated, or other preventive measures abandoned?

A statistical simulation is not deterministic like a mathematical model, or equation, which will usually give one answer when you feed in the relevant information. Instead, a stochastic model includes a process of randomisation to introduce an element of uncertainty into the situation, and many different outcomes may be possible.

How to design your own simulation

1 Observe the real situation, and decide which are the most important features or processes – such as, for example, the arrivals and service times in a queue, or the births and deaths for a population model.

2 Draw a flow diagram of the situation, identifying the main processes.

3 Which variables do you need to count or measure?

4 Can you model any of the measures by a statistical distribution? For example, can you model arrivals by a Poisson distribution?

5 If no distribution seems suitable, can you estimate probabilities by finding the relative frequencies for different outcomes?

6 If you have chosen a particular distribution as being a suitable model, which parameters do you need to estimate? These could include the mean for a Poisson

distribution, or the mean and standard deviation if a normal distribution is to be used to model the time taken, for example, for customers to be served.

7 Take a large sample of observations in order to estimate the parameters or relative frequencies required.

8 Set up the simulation by using a random number generator to randomise outcomes according to the probability model(s) used.

9 Compare the results from your simulation with a – preferably new – set of observations of the real situation.

10 Is your simulation realistic? If not, have you missed any important features in your model? Did you use the correct distribution?

11 If the simulation does seem to generate realistic results, you can use it to predict what might happen if changes are made.

Activity 6.1 A shop

The manager of a greengrocer's shop suspects that potential customers will walk away if the queue appears to be too long. He wants to find out how many people go away because the queue stretches outside the shop door, and he wants to determine the ideal number of shop assistants for the shop.

1 List two important processes in this situation, and draw a flow diagram for the situation to show how these processes relate to each other.

2 When the data are available, it may be possible to model either or both of the processes with a suitable statistical distribution. Suggest suitable distributions that you might be able to use to model each of the two processes. Give reasons for your suggestions.

3 Which parameters would you need to estimate for the distributions that you have suggested?

4 Now that you have decided what information you need to get from the data, you can decide what data you would need to collect in order to build your simulation, and on your collection method. Write a paragraph about how you could collect the data.

Using random numbers

Discrete probability distribution

First calculate the probability for each value of x. Table 6.1, for example, shows the probabilities for a Poisson distribution with $\mu = 1$. These probabilities have been converted to cumulative probabilities.

x	Probability $P(X = x)$	Cumulative probability $P(X \leq x)$
0	0.3679	0.3679
1	0.3679	0.7358
2	0.1839	0.9197
3	0.0613	0.9810
4	0.0153	0.9963
5	0.0031	0.9994
6	0.0005	0.9999
7	0.0001	1.0000

Table 6.1

Generate random numbers between 0 and 1, and take them as values of the cumulative probability, called the distribution function. Then, for each random number (y), you can find a corresponding value for x.

Random number (y)	Values of x
$0.0000 \leq y < 0.3679$	0
$0.3679 \leq y < 0.7358$	1
$0.7358 \leq y < 0.9197$	2
$0.9197 \leq y < 0.9810$	3
$0.9810 \leq y < 0.9963$	4
$0.9963 \leq y < 0.9994$	5
$0.9994 \leq y < 0.9999$	6
$y = 0.9999$	7

Table 6.2

Simulation of observations from a Poisson distribution $\mu = 1$

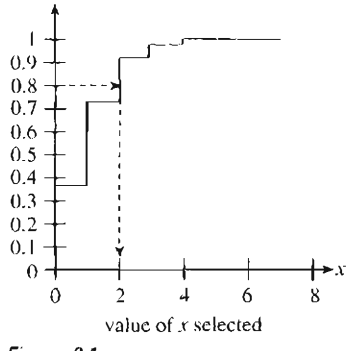


Figure 6.1

Notice that you need to use four-figure random numbers.

This graph shows the relationship between the random number (y) and the corresponding value of x.

For example, a random number 0.8000 selects an observation of $x = 2$ from a Poisson distribution with $\mu = 1$.

Activity 6.2 Discrete simulation

1 Draw up a scheme for selecting observations from a Poisson distribution with $\mu = 2$. Write a graphics calculator program to implement the scheme.

2 Show how you would use random numbers to select observations from a binomial distribution $B\left(10, \frac{1}{4}\right)$. Implement the scheme on your graphics calculator.

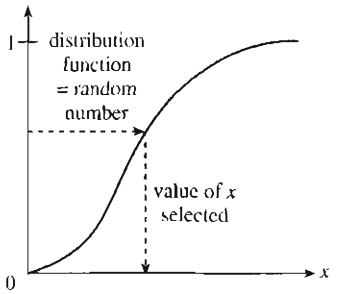


Figure 6.2

Continuous probability distribution

The method for sampling from a continuous distribution is similar to that for a discrete distribution, with random numbers corresponding to cumulative probabilities. The cumulative probabilities will form a continuous curve, as shown in Figure 6.2. This curve is a graph of the distribution function. Each random number generated will correspond to a unique value of x.

Simulating observations from a normal distribution

Statistical tables give the distribution function and the cumulative probabilities for the standard normal distribution $N(0,1)$. This makes it relatively simple to simulate observations from any normal distribution.

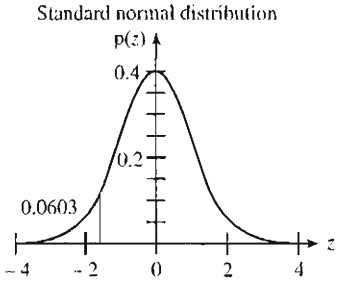


Figure 6.3

Example

Observations of customer service times taken at a supermarket check-out are found to follow a normal distribution with a mean of three minutes and standard deviation of 27 seconds. Here is the method for simulating an observation from that distribution.

a Generate a four-figure random number, for example 0.0630.
b Take this as the cumulative probability on the normal tables.

Which value of Z has an area to the left equal to 0.0630? This must be a negative value of Z. If 0.0630 is subtracted from 1, this gives 0.9370, the area to the left of the positive mirror image value of Z. From Table 1 in Appendix B, if $\phi(Z) = 0.9370$; then the magnitude of Z is 1.53, so, in this case, $Z = -1.53$.5
c Convert this to an observation from the normal distribution $N\left(3\,\text{min}, (27\,\text{sec})^2\right)$.

Then $Z = -1.53$, so $x = \mu - 1.53\sigma = 3\,\text{min} - 1.53 \times 27\,\text{sec} = 2\,\text{min}\ 19\,\text{sec}$ (to the nearest second).

Activity 6.3 Normal simulation

1 Use the random number 0.4286 to simulate an observation from the normal distribution with a mean of 2 min 30 sec and a standard deviation of 25 sec.

2 Look back at your flow chart for Activity 6.1, and simulate the queue at a greengrocer's shop. The average number of arrivals is two per minute. Customers take on average 2 min 30 sec to be served. There are three assistants. There is space to accommodate a queue of six people inside the shop. You will find this exercise easiest if you can use a spreadsheet.

3 How often do you find that people arrive when the queue already fills the shop?

There are no practice exercises for this chapter.

Reflecting on Chapter 6

What you should know

- how to use Poisson, binomial or normal distributions to build simulations.

7 Summaries and exercises

1 Introduction

Chapter summary

- how to recognise the four stages in the process of a statistical investigation (Activity 1.1)
- to understand the reasons why people collect data (Activity 1.2).

Practice exercises

There are no practice exercises for this chapter.

2 Collecting real data

Chapter summary

- how to recognise the different kinds of measurement scale: nominal, ordinal, interval and ratio; and some of the statistical techniques and models which may be used for different kinds of data (Activity 2.2 and the text before it.)
- how to recognise a true experiment, and how to identify the various kinds of experimental variable: explanatory, response, controlled and extraneous (Activities 2.3 and 2.4 and the preceding text.)
- understand some of the things that can go wrong in experiments (Activity 2.5), and some methods of experimental design which can help to overcome the problems (Activities 2.6 and 2.7)
- how to decide between three types of experimental design: independent samples, matched groups and before-and-after design (Activity 2.8)
- how to choose between various methods of sampling
  simple random sampling (Activity 2.9)
  systematic sampling (Activity 2.10 and preceding text)
  stratified sampling (Activity 2.11 and preceding text)
  quota sampling (Activity 2.12 and preceding text)
  cluster sampling (Activity 2.13 and preceding text)
- the importance of the response rate when you interpret the results of a survey (Activity 2.14)
- knowledge of some of the factors to consider when designing a questionnaire (Activities 2.16 and 2.17 and text between)

- the use of questionnaires in split polls (Activities 2.18 and 2.19).

Practice exercises

There are no practice exercises for this chapter.

3 Looking for differences

Chapter summary

- how to recognise the difference between non-parametric and parametric methods (text at end of chapter)
- how and when to use the Wilcoxon signed ranks test to analyse the results of an experiment using repeated measures or matched pairs (Activities 3.2 to 3.4 and surrounding text)
- how to calculate the probability distribution of T, the Wilcoxon signed ranks statistic (Activity 3.2)
- how and when to use the Mann-Whitney U test to analyse the results of an experiment with two independent samples (Activities 3.5 to 3.10 and surrounding text)
- how to calculate the probability distribution of U in the Mann-Whitney U test (Activity 3.6)
- how to use the tables for the Mann-Whitney U test and the Wilcoxon signed ranks test (Activity 3.11).

Practice exercises

1 Some students are given a course designed to improve their recall of facts in examinations. They are given a short memory test before and after the course. Their results in the test are shown in Table 7.1.

Student	A	B	C	D	E	F	G	H	I	J
Score before	10	9	8	7	6	12	9	14	11	10
Score after	11	7	10	10	9	11	13	14	12	9

Table 7.1

Do the results show conclusive evidence that the course is effective? Justify whether you need a one- or two-tailed test, and clearly state your assumptions.

2 Table 7.2 shows the barometric pressure in millibars at two meteorological observatories A and B, taken at random times.

A	1002	976	998	1026	1034	1005	972
B	970	984	986	1003	987	1003	

Table 7.2

Do these figures give significant evidence of a difference between the barometric pressures at A and B?

3 A physiotherapist has investigated the effectiveness of a new exercise for use in the rehabilitation of patients who have suffered spinal injury. The patient's fitness was measured by recording the time taken to complete the exercise. Ten patients tried the exercise daily for two weeks and Table 7.3 shows the times taken (in seconds) to complete the exercise on the first day and the last day.

Patient	A	B	C	D	E	F	G	H	I	J
Time on first day	27	45	51	33	49	22	57	35	41	25
Time on last day	22	37	45	26	50	20	60	35	42	23

Table 7.3

What does this experiment indicate about the effectiveness of the exercise programme?

4 Fitting a model to the data

Chapter summary

- how to use the χ^2 distribution to test the fit of a given model (Activities 4.3 and 4.4, and the surrounding text, and Activity 4.7)
- how to fit a binomial distribution to a frequency distribution, and how to test the fit of your model (Activities 4.5 and 4.8 to 4.10, and the text following 4.10)
- how to fit a normal distribution to a frequency distribution, and how to test the fit of your model (Activity 4.11)
- how to explore the possibility of fitting a normal distribution to a frequency distribution, using normal probability graph paper (Activities 4.12 to 4.15, and the text following 4.15)
- how to use the χ^2 distribution to test the association between two nominal variables (Activities 4.16 and 4.17).

Practice exercises

1 a In a breeding experiment it is expected that ducks would hatch in the ratio of one duck with a white front to every three without. The experimenter observes 96 ducks hatch, of which 18 have white fronts. Do these results seem to be consistent with the expectations?

b In one week a public library records the number of books borrowed on each day. Is there reason to believe that more books are borrowed on one weekday than on any other?

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
No. of books issued	190	278	228	252	225	273

Table 7.4

2 A student rolls a dice 1200 times and records on each roll whether the score is even or odd, and large (4, 5, 6) or small (1, 2, 3). Table 7.5 shows the frequencies he observes.

	Large	Small
Even	375	183
Odd	212	430

Table 7.5



Carry out a chi-squared test of the fairness of the dice.

3 Refer to the data in Table 7.6 to answer the following questions.

Value x	0	1	2	3	4	5	6	7	8
Observed frequency	2	4	10	15	19	12	8	7	1

Table 7.6

a Use a chi-squared test to test whether the values of x can be treated as observations from a binomial distribution where $n = 8$.
b Assume that the values of x are observations of a continuous variable recorded to the nearest whole number, and fit a normal distribution to the data. Use a chi-squared test to test the fit.

5 The Poisson distribution

Chapter summary

- when and how to use the Poisson distribution to model the occurrence of random events (Activities 5.2 and 5.5, and the text following Activity 5.5)
- how the Poisson distribution is related to the binomial distribution (Activity 5.3)
- how to use the Poisson distribution as an approximation to the binomial distribution when n is large and p is small (Activities 5.1 and 5.4)
- how to use the Poisson distribution as a model in hypothesis testing (Activities 5.7 and 5.8)
- how the shape of the Poisson distribution changes as μ increases (Activity 5.9)
- how to fit a Poisson distribution to a frequency distribution and how to test the fit of your model using a χ^2 test (Activity 5.10)
- how to use Poisson probability graph paper to explore the possibility of fitting a Poisson distribution to a frequency distribution (Activity 5.11).

Practice exercises

1 A large manufacturing company has kept records over many years of the number of employees who have died at work on the company's business. The average number of deaths per year is 0.7. Two years ago there were four deaths in 12 months, and last year there were two. Is this evidence that the underlying death rate has increased?

2 In 1898 a pamphlet was published which gave details of the number of men killed by horse kicks each year for each of the 14 corps in the Prussian army from 1875 to 1894. This information, as shown in Table 7.7, has become one of the most often quoted examples of a Poisson distribution.

Number of men killed	0	1	2	3	4	5 or more	Total
Frequency	144	91	32	11	2	0	280

Table 7.7

a Explain how the total frequency matches the information given about the source of the data.

b Fit a Poisson distribution to the data, and test the 'goodness of fit'.

6 Simulations

Chapter summary

- how to design your own simulation (see the text before Activity 6.1)
- how to use random numbers generated on a calculator or a computer to simulate values from a discrete probability distribution such as a Poisson or binomial distribution (see Activity 6.2), or from a continuous distribution such as a normal distribution (see Activity 6.3).

Practice exercises

There are no practice exercises for this chapter.

8 Hints

Activity 2.1, page 9

1 **a** In seeking the opinions of voters, it is not possible to interview them all (as in a census). A sample or smaller group has to be selected. First the population, or group of people referred to, must be clearly defined.

b Does the research question refer to all adults now over 18, or to people who will be over 18 by the next general election, or to some other population? Not all people over 18 will be eligible or registered to vote. So does it refer to all potential voters – that is to say, all adults over 18 who are on the electoral register? Or should only those people who actually intend to vote be considered?

2 To measure the effect of the fertiliser on the growth of plants requires an experiment. How would you design the experiment?

3 To answer the question 'Does anxiety improve people's performance in examinations?' also requires an experiment. How would you design a suitable experiment?

4 Do people in the UK favour proportional representation? Does this question refer to the same population as the first question, or should a wider group be considered?

Activity 2.3, page 13

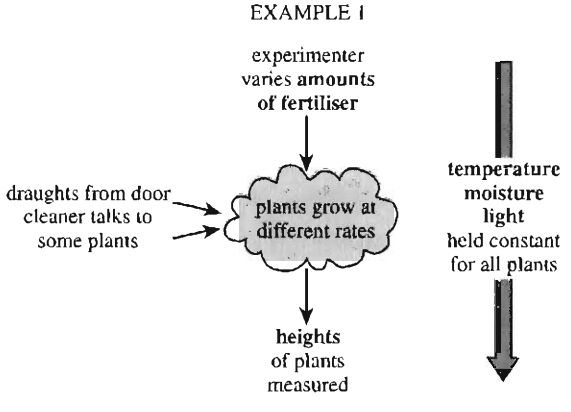


Figure 8.1

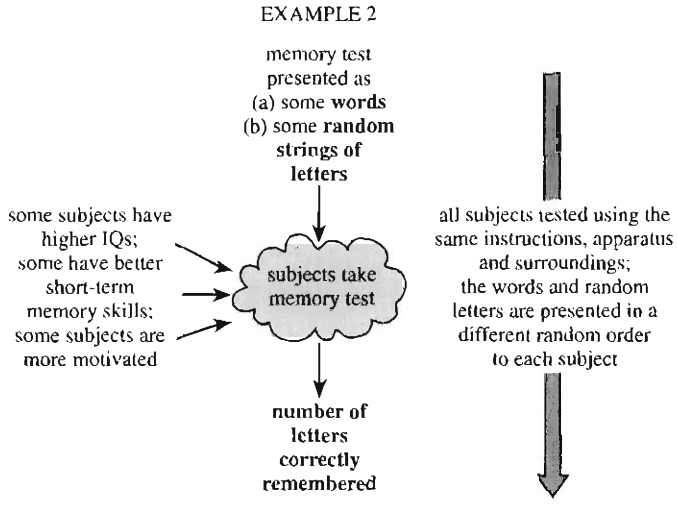


Figure 8.2

Activity 2.11, page 22

1 You need to decide on the strata for your sample. At the least you will need to have a strata of lower floors separate from the others, and you might decide that you want a sample from the penthouse flats separately too.

Activity 2.14, page 25

2 The problem with non-responses is that they may not be distributed across the whole population. In this case, consider what reasons there might be for a non-response. The reasons are likely to be different depending on whether you are collecting the data by a postal enquiry or by a personal call. You need to weigh up how these reasons might distort the results that you obtain from your sample.

Activity 3.1, page 36

9 Remember to ignore the zero differences.

Activity 3.2, page 37

2 **a** Suppose that there were only two ranks. Each could be positive or negative, so there would be $2 \times 2 = 4$ possible arrangements of pluses and minuses. If there were three ranks, each could be either plus or minus, giving $2 \times 2 \times 2 = 8$ possible arrangements of pluses and minuses. Now try generalising to n ranks.

d You can answer this easily by looking at the table that you used to answer question **2b**, and imagining each plus changed to a minus, and each minus changed to a plus.

Activity 3.5, page 42

3 This counting is easily done if you use the diagrams that you drew for question **2**.

4 Look at the values of U_1 and U_2 for all your sets of data. Make a note of any similarities and differences that you can see. What is the smallest possible value that U_1 can take? What is the largest? What can you say about the sum of U_1 and U_2?

Activity 3.6, page 43

1 a You will need three counters, one for the N's, one for the S's and one for U_1. Every time you get an N, add 1 to the counter for the N's, and add the current value of the counter for the S's to U_1.

Write your algorithm so that if you obtain three N's, the procedure stops and outputs U_1. Any remaining spaces will be assumed to be filled with S's.

b *ProbU* will need two loops: one to set up the frequency table for the possible values of U from 1 to 9, and the second to call *FindU* and generate values of U.

3 b For each value of U, count the number of arrangements of three N's and three S's which make it, and multiply by the probability that you found in question **3a**.

Activity 4.1, page 57

1 Calculate the mean number of fives and sixes in each throw of 12 dice. You can use a graphics calculator or a spreadsheet to do this. Then the proportion of fives and sixes is the mean number of fives and sixes divided by 12.

2 The probability of scoring five or six on each dice is $\frac{1}{3}$.
The number of dice thrown is 12 each time.
The dice can be expected to fall independently.

3

Weldon

```
Input M                                   {M is a positive integer, the number of throws of 12 dice}
  (2,13) → dim X                          {set up X as a 2 × 13 array or matrix}
  for J = 1 to 12
    J → X(1, J + 1)                       { X(1, J) is the number of dice scoring five or six }
    0 → X(2, J + 1)                       { X(2, J) is the frequency of getting J dice scoring five or six}
  next
  for N = 1 to M
    0 → D                                 {D will count the number of dice scoring five or six}
    for J = 1 to 12
      D + int(rand + (1/3)) → D           {If the dice thrown scores five or six, adds 1 to D}
    next
    X(2, D + 1) + 1 → X(2, D + 1)
  next
  display X
Output X, the frequency table of fives and sixes in 12 throws of a dice.
```

Activity 4.2, page 58

3 The iterative formula is $P(r+1) = \frac{(n-r)}{(r+1)} \frac{p}{q} P(r)$.

4 The expected frequency of 0 fives or sixes in 12 dice is $7006 \times P(0)$.

Activity 4.5, page 62

3 Look at your answer to question **1** of Activity 4.1.

7 Think carefully about the number of constraints that you would have if you were to make up a set of frequencies such that the total frequency was 7006, and the proportion of successes equal to the estimate that you found in question **3**. How many free choices would you be able to make?

10 The analysis in questions **1** to **8** assumes that the probability of a five or a six is the same for all the 12 dice. Thus either all 12 dice are fair for the analysis in Activity 4.4, or they are all biased in the same way for this activity. How could you analyse the situation where only some of the 12 dice are biased?

Activity 4.6, page 63

2 You are expecting a uniform distribution with $p = \frac{1}{6}$ for an unbiased dice. The alternative hypothesis is simply that the scores are not equally likely.

7 Run your simulation of 120 throws of a fair dice, and calculate the value of $\sum \frac{(O-E)^2}{E}$ each time. How many times out of 100 is this value significant? A significant value at the 5% level is one greater than the value of $\chi^2_\nu(5\%)$, where ν is the number of degrees of freedom.

Activity 4.8, page 66

2 There are eight different colours possible for Smarties, and a useful modelling assumption is that they are all equally likely.

Each Smartie is either red or not red, and there are assumed to be 16 Smarties in each packet.

4 Expected frequency = number of trials × probability.

8 If you reject the null hypothesis, you will need to consider which of the assumptions made when you use a binomial model might be invalid in this situation. Is the number of trials fixed, or is the probability of 'success' not constant, or are the trials not independent?

Activity 4.9, page 68

6 There is only one class with an expected frequency less than 5. You may like to compare the results of leaving this as a separate class, keeping the degrees of freedom as 5, or combining classes but reducing the degrees of freedom.

10 From the tables of χ^2, the critical value for χ^2 with $\nu = 5$ is 11.07, whereas that for $\nu = 3$ is only 7.81.

Activity 4.10, page 68

3 **a** Calculate the mean number of pins landing point-up out of ten, and the probability of a pin landing point-up, for:

- each student
- all the students together

- all the students except student G.

Fit a binomial distribution to the data from all the students, and use a χ^2 test to test how well it fits the observed frequencies. Use the method of Activity 4.9.

b Do the same for the observed frequencies, excluding student G's results.

Activity 4.12, page 72

6 The area to the left of $Z = 1$ on a standard normal distribution is 0.84. Thus 0.84 is the probability of observing a value less than the mean plus one standard deviation.

Activity 4.13, page 73

1 Make a cumulative relative frequency table such as the one that you made in questions **1** and **2** of Activity 4.12. Because you need to treat the data as continuous, the upper limit of the class intervals are 53.5, 57.5, 61.5, etc. Then you can plot the cumulative relative frequency against the upper limit for each class interval, using your normal probability graph paper.

Activity 4.16, page 77

1 Null hypothesis: the proportion of smokers who are vegetarians is the same as the proportion of all people who are vegetarians.

Alternative hypothesis: the proportion of smokers who are vegetarians is different from the proportion of all people who are vegetarians.

3 The expected number of smokers who are vegetarian is $\frac{32}{95} \times 26 = 8.758 = 8.8$ to one decimal place.

In this way the 26 smokers can be split in the ratio 32:63 between vegetarian and non-vegetarian. The 63 non-smokers must also be split in the ratio 32:63.

A useful algorithm to calculate the expected frequencies is:

$$\text{expected frequency} = \frac{\text{column total} \times \text{row total}}{\text{grand total}}.$$

5 Try this:

	Expected frequencies		
	Vegetarian	Non-vegetarian	Totals
Smokers	8.8		26
Non-smokers			69
Totals	32	63	95

Table 8.1

Can you work out the other expected frequencies? (By subtraction.)

If you fix any one of the frequencies, can you work out all the other frequencies?

Activity 5.2, page 85

2 Your frequency table should look like this:

Number of arrivals	Frequency
0	
1	
2	
3	
4	
etc.	
Total	

Table 8.2

Activity 5.4, page 87

4 The number of degrees of freedom will be the number of cells minus one, since the only constraint in each of the models is that the total expected frequencies have to match the total observed frequencies.

Activity 5.5, page 87

3 Multiply each Poisson probability by your total frequency (that is, the number of time intervals in your survey) to give the expected frequencies.

5 Remember to group the frequencies. There are two constraints in this case. You have used the observed frequencies to estimate the value of the mean (μ) for your theoretical model, and the total expected frequencies must be the same as the total observed frequencies.

6 Consider whether, in the situation that you have observed, the arrivals occur uniformly. It may be that the rate of arrivals is affected by some outside influence. In the example of traffic arriving at a car park, there may be a set of traffic lights nearby which interferes with the flow of traffic arriving at the car park.

You should also consider whether the arrivals are occurring independently. If, for example, there is a queue for the car park, then the arrivals will pass through the entrance at fairly regular intervals. If you are recording the arrival of cars to join the queue for the entrance, then the presence of a queue might influence some drivers not to attempt to enter the car park at all, and therefore affect the rate of arrivals.

Activity 5.6, page 89

5 Let x be the minimum stock needed at the start of each week. Then the probability of not having enough light bulbs in a week is the probability that more than x bulbs need replacing. You might find it helpful to use a spreadsheet to answer the question now.

Activity 5.8, page 90

1 What is the usual average for six months?

2 Calculate the probability of receiving four or more calls.

4 What is the average number of letters for one day?

Activity 5.9, page 90

1 $P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$

$P(X = x - 1) = \frac{\mu^{x-1}}{(x-1)!} e^{-\mu}$

$\frac{P(X = x)}{P(X = x - 1)} = \frac{\mu}{x}$

Practice exercises

Chapter 4, page 103

2 Notice that the null hypothesis is that the dice is fair, and this gives the theoretical probabilities, so you do not need to estimate the proportions from the data. Therefore the only constraint on the expected frequencies is that the total should be 1200.

9 Answers

Activity 1.1, page 1

1 Define the problem: Why is the incidence of lung cancer increasing?

Collect data, and analyse.

Interpret the data: smoking is associated with lung cancer. Notice that there is not yet evidence that one causes the other.

Redefine the problem: does smoking cause lung cancer? Is smoking associated with other forms of cancer? Is smoking associated with other diseases?

Collect more data. Notice that the method of data collection may need to be different here from the method used previously.

2 The first claim may stand on its own, but you might need to know how 'smokers' are defined, and what distinguishes them from 'non-smokers'. For example, is a person who last smoked three months ago a smoker or a non-smoker? What about the person who smokes very occasionally, less than once a month? A research question arising out of this might be: what is the effect when heavy smokers (smoking more than 20 cigarettes per day) are compared with non-smokers? What is the effect when long-term smokers (regular smokers for more than 30 years) are compared with non-smokers? Does the effect change if, instead of non-smokers, you compare with people who have never smoked?

For the second claim you would need to know how the number of deaths caused by tobacco smoking has been estimated. At first sight, it is not easy to see how to define that a death is caused by tobacco smoking.

Activity 2.1, page 9

1 **a** Survey
b The people asked
c All potential voters in the next general election? (See the hint)
d There may be bias in your sampling process.

2 **a** Experiment
b The plants in the experiment.
c All plants which might be dosed with the fertiliser? (See the hint)
d There may be other effects arising in the experiment which have not been controlled, or there may be extraneous variables. (See after Activity 2.2)

3 **a** Experiment
b The people in the experiment
c All people sitting examinations? (See the hint)
d There may be other effects arising in the experiment which have not been controlled, or there may be extraneous variables. (See after Activity 2.2)

4 **a** Survey
b The people asked
c All potential voters in the next general election, maybe? (See the hint)
d There may be bias in your sampling process.

Activity 2.2, page 12

TOWNS dataset:

Column 1	Ratio
Column 2	Ratio
Column 3	Ratio
Column 4	Ratio
Column 5	Ratio
Column 6	Ratio
Column 7	Nominal
Column 8	Ordinal
Column 9	Nominal
Column 10	Possibly only ordinal since it may not form an interval scale.
Column 11	Nominal

HEIGHT dataset

Column 1	Ratio
Column 2	Ordinal
Column 3	Ratio
Column 4	Ratio
Column 5	Ordinal
Column 6	Ratio
Column 7	Nominal
Column 8	Ratio
Column 9	Nominal

Activity 2.3, page 13

Example 1
The object of the experiment is to discover how the growth of the plants is affected by different amounts of fertiliser. Therefore the response variable is the height of the plant after a defined time. The experimenter is hoping to explain this by the amount of fertiliser applied to the plants, so this is the explanatory variable. Many variables, such as temperature, light and moisture, can be controlled to some extent by conducting the experiment under controlled conditions. But there may be other variables that are either impossible to control, or are not known to the experimenter. Examples of these extraneous variables may be draughts from the door, or perhaps the cleaner has got into the habit of talking to some of the plants each day!

Example 2
Each subject in this experiment is given a memory test consisting of several strings of letters. The response variable is the number of letters correctly remembered by each subject. Some of the strings are random strings, and others are words that will have some meaning to the subjects. This distinction is the explanatory variable, since the experimenter suspects that it may explain the response variable. The experimenter will try to ensure that each subject is tested under identical conditions, using the same apparatus and in the same surroundings. These factors are controlled variables. There is a possibility that subjects may with practice get better at the tasks in the memory test. To control this practice effect, the words and random letters are presented to each subject in a different random order.

Activity 2.4, page 14

1 a Explanatory: the presence or absence of fire retardant spray
Response: the time taken to burn
Controlled: the size of the fabric sample and, since a fume cupboard is used, the amount of oxygen supplied and the background temperature will be held constant; also the flame should be constant.
Extraneous: none

b Explanatory: the series of practice IQ tests
Response: the improvement, or change, in IQ scores between first and sixth test
Controlled: the same people used throughout the experiment
Extraneous: some people may do other practice activities on their own, that is to say other IQ tests or puzzles. Some will have had previous experience of tests and may already have improved their scores to the limit. (There may be a limit to the amount of improvement which can be brought about as a result of practice.) People's test scores will be affected by whether they feel ill or tired at the time of the test. At the end of six weeks of tests they may feel bored, and lose motivation.

2 a Alcohol has not been intentionally administered by an experimenter. The drivers' performance before and after alcohol cannot be compared. We have no idea whether these drivers would have been involved in accidents had they not been drinking. There is no control of the situation by an experimenter. Do drivers who drink have other features which make them more prone to having accidents? (Are they younger or less experienced?)

b There is no control of the subjects entering the two dental practices. It may be that parents who are particularly keen on preventive dentistry actively choose to take their children to the first dentist. These parents may give their children more encouragement with other preventive measures; for example, brushing teeth more often, more thoroughly, eating less sugar, and so on.

3 Subjects should be presented with three samples: two the same and one different. The order in which the samples are presented should be randomised, and some subjects will receive two of tap water and one of bottled water, whilst others will receive two of bottled water and one of tap water. There should be no background smells or noises to distract subjects, who should all be given the same instructions. The temperature of all samples should be the same.

Activity 2.5, page 14

1 a The effect of the exercise programme may be confounded with the effect of the experiment itself.

Participants will be interviewed individually at the start and end of the programme and will also have increased social contact with each other during the exercise sessions. These social aspects may lead to an improved self-esteem irrespective of any improvement in physical capabilities.

The problem can be avoided by using a control group, as much like the experimental group as possible. The control group will need to have weekly contact without the exercise, perhaps a coffee and chat group.

b There is a possibility of a placebo effect here. A second control group is required, each of whom receives a dummy pill, and ideally a third group who receive no pills at all.

Secondly, there may be an experimenter effect. If the teacher knows which children have received the genuine pills she may be expecting to see improvements in performance from those children. This positive expectation may lead to those children's performance improving. The control here is to ensure that the teacher does not know which children are which.

Some children's performance will improve anyway due to maturation. The control group will enable this to be seen.

c You might expect the second run to be slower than the first if the athletes are tired, or you might think they will be faster the second time because they have 'warmed up'.

One way to control this would be to insist on an adequate rest period between runs, but you might also randomise the order in which the shoes are worn by each athlete. In this way any difference due to order will be averaged out between the two types of shoe over the whole sample.

Activity 2.7, page 17

1 **a** Explanatory variable: new method of teaching reading. Response variables: reading scores of pupils. Confounding variables: teachers are volunteers who will be well motivated and committed to the new method. Their expectations may well be communicated to the pupils who will respond positively to their teachers' enthusiasm. We do not know whether children will have come into school already able to read – no mention is made of testing them at the beginning of the experiment. Also some children may receive extra tuition at home – parents may react positively, or possibly negatively, to their children taking part in an experiment.

b Explanatory variable: use of vitamin pills. Response variable: amount of absence from lessons. Children at boarding school may find it difficult to take time off school unless they are genuinely ill! Confounding variables: pupils in year 7 starting a new school are presumably under more stress in a new school than they will be in year 8. Stress is an important factor in causing illness. The weather and amount of background illness in the community will vary from one year to the next, so the experiment has not been conducted under constant conditions. Some pupils taking the vitamin pills may experience a placebo effect and feel better because of this. A control group should be used taking dummy pills.

c Explanatory variable: is the test a number test or a letter test? Response variable: students' test scores. Confounding variable: all students are taking the number test second so there is likely to be a practice effect. The order of the tests should have been randomised. Not all the students selected may be particularly good at maths!

d Explanatory variable: course of educational videos. Response variable: difference in pupils' test scores before and after course of videos. Confounding variables: pupils may be exposed to other influences such as TV documentaries, news of an ecological disaster etc. A control group should have been tested in the same way, without the video course in between.

2 No; it is not a correct interpretation. The placebo effect does bring about genuine improvements in physical symptoms. Rather like faith healing, a positive attitude does bring about real physical recovery in some cases. The fact that no treatment has been given does not mean that the patient was not really ill.

Activity 2.8, page 19

First statement: Subjects would be given a chance to practice on a simulator before taking a test without alcohol. Then they would be asked to drink a controlled amount of alcohol and time would be allowed for this to be absorbed into the bloodstream. The same subjects could be tested several times, with increasing amounts of alcohol.

a Driving skill is not affected by the level of alcohol in the blood.
b Driving skill is affected, adversely, by the level of alcohol in the blood.

c Amount of alcohol drunk
d Number of mistakes made in driving simulator, compared to score before any alcohol is drunk
e Same subjects before and after drinking
f Control the amount of alcohol taken and the time allowed between drinking and re-testing. Also control or at least record the time and size of the subjects' last meals – drinking on an empty stomach may have a greater effect than after a heavy meal. The amount of practice may be less important (students will have varying amounts of experience in driving) it is the *change* in performance that is being tested.
g Difficult to randomise order in which they take the tests with and without alcohol!

Second statement: Subjects are to be shown cards, or slides, each with two letters printed on,

e.g. **CD** **dD** **de** etc. The following six combinations are possible

(i) two letters the same, both upper case
(ii) two letters the same, both lower case
(iii) two letters the same, one upper case, one lower
(iv) two letters different, both upper case
(v) two letters different, both lower case
(vi) two letters different, one upper case, one lower.

The order of these should be randomised. At each testing (card) the subject must say 'same' or 'different'. They can be timed using a stop-watch, and also the number of mistakes counted – most likely to be in category (iii).

a Subjects can identify when two letters are the same, irrespective of whether they are written in the same case or not.
b Two letters can be identified as having the same name more rapidly if they are written identically.
c Whether the letters are written in the same case or not
d Time taken to give answer or number of mistakes made
e Same subjects throughout
f Controlling colour of ink; size of cards; style of writing; subjects' surroundings, background noise etc.; instructions to subjects
g Randomise order of cards, and combinations of letters.

Third statement: For a randomly selected group of subjects, ask them to weigh themselves last thing at night and first thing in the morning (before visiting the bathroom!) The scales used need to measure weight correct to 0.1 kg.

a There is no difference in people's weights last thing at night and first thing in the morning.
b People weigh more last thing at night than they do first thing in the morning.
c Lack of food during the night
d Change in weight
e Same subjects in both conditions
f Ask subjects not to eat or drink during the night or visit the lavatory! Ensure that identical clothing is worn at both times.
g None

Activity 2.9, page 20

1 Expensive and difficult to find the members of a simple random sample to ask them the question.

2 Identify different sub-groups of the population by age, social class, ethnic origin etc that may (or may not) have different voting intentions and ensure that each sub-group is represented in the survey.

Activity 2.10, page 21

1 Yes

2 Yes

3 Approximately 24/365, assuming births are uniformly distributed throughout the year

Activity 2.11, page 22

1 Many designs are possible.
One possibility is:
Floors 1–2: sample 6 flats
total children on these floors is $n \times 84 \div 6$
Floors 3–5: sample 9 flats
total children on these floors is $n \times 126 \div 9$
Floors 6–10: sample 14 flats
total children on these floors is $n \times 210 \div 14$
Floors 11–17: sample 20 flats
total children on these floors is $n \times 294 \div 20$
Floor 18: sample 1 flat, estimated total children on this floors is $n \times 18$
where n is the total number of children under 16 in the families sampled.

This gives a total sample of 50 families, and ensures that there is always a fair proportion of families from the lower floors.

Activity 2.12, page 23

1 A simple random sample, systematic or stratified sample would require the researcher to have a complete list of the target population. Such a list would be difficult to come by in this case since the young people in question are not found in any one place. Some are in school, some in college, some in employment and some unemployed.

Even if such a list could be obtained or constructed, the young people selected for the sample would be located randomly all over the country, and so would be difficult to contact. Probably the only way in practice to contact them would be to send out the questionnaire without prior warning to all members of the sample. This would be likely to result in a very high rate of non-response.

2 The young people in the sample will tend to be those who are relatively easy to contact. Young people who are homeless or ill, for example, may be under-represented.

Activity 2.13, page 23

2 There are 63 aged 34 and under, and 137 aged 35 or more. About 0.3 are in the younger age group.

3 Sample 6 from age 34 or less, and 14 from age 35 or more.

9 You will find that the results for the stratified and quota samples are generally less variable than the results for the simple random sample or the systematic sample.

Activity 2.14, page 25

1 Older people and those with no dependent children may be more often away from home, less often home, and so more difficult to contact. Older people may be ill in hospital, or away visiting relatives or on long holidays.

Particular caution should be exercised when drawing conclusions related to the groups of people who are under-represented.

Activity 2.16, page 26

1 c It is important to keep sentences and questions short, and to use simple language. A maximum of about 20 words is a good guide for the length of a question. Do not use a long word if a short word will do. If you want members of the general public to complete your questionnaire, you should remember that the reading age of most tabloid newspapers is only about 11 years. The skills needed to read and fill in your questionnaire may be ones which people use only rarely. Questions need to be framed in such a way as to get responses that are easy to analyse. Ask "Do you prefer X, Y, Z, other?" rather than "What is your favourite...?" This also avoids a literacy problem, as respondants do not need to spell their responses.

2 a It is not clear what is meant by 'average'.

b Subjects will feel it is not socially acceptable to say 'no'.

c Too general. People will have different ideas about what it means to be a healthy person!

d Most people will have had a headache at some time. This question needs more introduction to enable people to judge the kind of headache you are asking about. A list of categories of severity of headache could be provided for people to tick.

e Two questions.

f Leading question.

g Two negatives – *not* and *non*-medicated. Try 'Do you prefer to use a medicated shampoo?'

h Too difficult for people to give you an accurate answer! If you really want to know, use a pedometer (that is to say, measure it).

i Too difficult to remember accurately, depending on the length of the time gap.

j This is too difficult, since most people will not have the information needed to answer.

k Not many people will understand this question! Vocabulary too difficult.

Activity 3.1, page 36

1

List A		List B
2	0	0 3 4 3 4
7 5 5 9	0	5 6 8 6 9 5
1 2 4 4 0 0	1	3
	1	
1	2	

$n = 12$

	List A		List B	
M	10		5	
Q	6	13	3.5	7
X	2	21	0	13

2 These analyses show that subjects scored more highly on list A, with a median score of 10 compared to 5 on list B. Both the range and the inter-quartile range are greater for test A scores than test B scores, indicating that subjects' scores were actually more variable on test A also. These analyses cannot show the difference in scores achieved by individual subjects on list A compared to list B.

3

Differences: 9, 5, 4, 6, −1, 11, 12, 0, −2, −1, 8, 3
In order: −2, −1, −1, 0, 3, 4, 5, 6, 8, 9, 11, 12

4

	Differences	
M	4.5	
Q	−0.5	8.5
X	−2	12

5 This shows that a few people, about one quarter, scored better on list B than list A. However, most people scored better on list A with a median difference of 4.5 words. The greatest difference was 12.

6 Null hypothesis: any differences in scores between list A and list B are completely random. There is no systematic difference between subjects' scores in the two lists.

Alternative hypothesis: subjects will score more highly on list A because the words in that list are meaningful pairs, whereas the words on list B are not.

7 The differences between the list A and list B scores are assumed to be independent.

8 There are eleven non-zero differences; eight are positive and three are negative.

9 $X \sim B\left(11, \frac{1}{2}\right)$

x	$P(X = x)$	
0	$\left(\frac{1}{2}\right)^{11}$	0.000 49
1	$11 \times \left(\frac{1}{2}\right)^{11}$	0.005 37
2	$\frac{11 \times 10}{2 \times 1}\left(\frac{1}{2}\right)^{11}$	0.026 86
3	$\frac{11 \times 10 \times 9}{3 \times 2 \times 1}\left(\frac{1}{2}\right)^{11}$	0.080 57
	Total probability	0.113 29

This probability is too large to reject the null hypothesis. Using the sign test, which focuses only on the number of positive and negative differences, the conclusion is that any differences between subjects' scores on list A and list B can be attributed to random variation.

Activity 3.2, page 37

1 a

Difference: 9, 5, 4, 6, −1, 11, 12, 0, −2, −1, 8, 3

b

Abs. diff: 9, 5, 4, 6, 1, 11, 12, 0, 2, 1, 8, 3

c

Rank: 9, 6, 5, 7, 1.5, 10, 11, −, 3, 1.5, 8, 4
Sign: + + + + − + + − − + +

d Total of all the ranks awarded to positive differences = 60

e Total of all the ranks awarded to negative differences = 6

2 a $2^{11} = 2048$

b 13

c If the null hypothesis is true, then $P(T \leq 6) = 13/2048 = 0.0063$. Therefore this result is extremely unlikely, and you conclude that subjects score more highly on list A.

d There are 13 ways of getting a total of 6 or less for the sum of ranks awarded to positive differences.

e Using a two-tailed test, $P(T \leq 6) = 26/2048 = 0.0127$. Again, this is an unlikely result, and is significant at the 5% level.

f The Wilcoxon signed ranks test takes more information into account than the sign test does. It is therefore more likely to give a significant result from a small set of data.

Activity 3.3, page 39

1 a Null hypothesis: There is no difference between subjects' weights at night and in the morning.

b Alternative hypothesis:
either:
There is a difference between subjects' weights at night and in the morning. They will weigh less first thing in the morning. (One-tailed test)

or

There is a difference between subjects' weights at night and in the morning. (Two-tailed test) [This alternative is not incorrect.]

d

Evening	Morning	Difference	Rank of absolute difference	
28.0	27.4	0.6	6	(+)
35.0	35.0	0	ignore	
54.5	54.0	0.5	4	(+)
68.0	68.0	0	ignore	
70.5	70.0	0.5	4	(+)
42.5	43.0	−0.5	4	(−)
51.0	50.0	1.0	10	(+)
55.0	54.2	0.8	9	(+)
50.0	49.9	0.1	1	(+)
63.6	63.2	0.4	2	(+)
69.1	68.4	0.7	7.5	(+)
31.0	30.3	0.7	7.5	(+)

Note that there are three values at 0.5, some positive and some negative. Therefore they must all be ranked at 4, the average of the positions (3, 4 and 5), which they would have filled had they had different values.

Total of positive ranks = 51

Total of negative ranks = 4

Ranks										T
1	2	3	4	5	6	7	8	9	10	
+	+	+	+	+	+	+	+	+	+	0
−	+	+	+	+	+	+	+	+	+	1
+	−	+	+	+	+	+	+	+	+	2
−	−	+	+	+	+	+	+	+	+	3
+	+	−	+	+	+	+	+	+	+	4
+	+	+	−	+	+	+	+	+	+	4
+	+	+	+	−	+	+	+	+	+	4

There are $2^{10} = 1024$ different orders possible.
So $P(T \leq 4) = \frac{7}{1024}$

On a one-tailed test this is a tail probability of 0.0068 and on a two-tailed test, 0.0137.

Whether you use a one-tailed or a two-tailed test, this result is significant.

e People do weigh significantly less in the morning than they do last thing at night.

2 a Null hypothesis: students have times which are equally fast for breaststroke and butterfly when swimming 100 metres.

b Alternative hypothesis:

either

Students are faster at swimming butterfly than breaststroke (one-tailed test).

or

Students' times for swimming butterfly and breaststroke are not the same (two-tailed test).

Breast stroke	Butterfly	Difference	Rank of absolute difference	
92	90	2	2	(+)
80	83	−3	3.5	(−)
102	98	4	5.5	(+)
105	106	−1	1	(−)
81	75	6	7.5	(+)
96	93	3	3.5	(+)
88	84	4	5.5	(+)
76	70	6	7.5	(+)
100	91	9	9	(+)
97	84	13	10	(+)

The total for the positive ranks = 50.5

The total for the negative ranks = 4.5

$T = 4.5$.

Ranks										T
1	2	3	4	5	6	7	8	9	10	
+	+	+	+	+	+	+	+	+	+	0
−	+	+	+	+	+	+	+	+	+	1
+	−	+	+	+	+	+	+	+	+	2
−	−	+	+	+	+	+	+	+	+	3
+	+	−	+	+	+	+	+	+	+	3.5
+	+	+	−	+	+	+	+	+	+	3.5
−	+	−	+	+	+	+	+	+	+	4.5
−	+	+	−	+	+	+	+	+	+	4.5

There are $2^{10} = 1024$ different possible orders.

The probability of $T \leq 4.5$ is

one-tailed test = 0.0078

two-tailed test = 0.0156.

This result is significant on both a one-tailed and a two-tailed test.

e The students are faster at swimming 100 m using butterfly.

3 **a** Null hypothesis: the athletes run 400 m equally fast at sea level and at high altitude.
b Alternative hypothesis: athletes run faster at sea level than at high altitude.
c One-tailed test

Sea level	High altitude	Difference	Rank of absolute difference
47.5	47.2	0.3	1 (+)
48.9	47.8	1.1	3 (+)
47.9	48.5	–0.6	2 (–)
48.4	50.5	–2.1	6 (–)
51.2	54.6	–3.4	7 (–)
48.0	49.8	–1.8	4 (–)
50.2	52.1	–1.9	5 (–)
50.8	54.9	–4.1	8 (–)

There are $2^8 = 256$ possible orders. Total of positive ranks = 4; total of negative ranks = 32. $T = 4$.

Ranks								T
1	2	3	4	5	6	7	8	
–	–	–	–	–	–	–	–	0
+	–	–	–	–	–	–	–	1
–	+	–	–	–	–	–	–	2
–	–	+	–	–	–	–	–	3
+	+	–	–	–	–	–	–	3
–	–	–	+	–	–	–	–	4
+	–	+	–	–	–	–	–	4

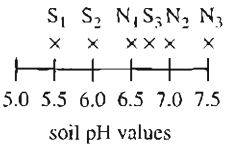
$P(T \leq 4) = \frac{7}{256} = 0.0273$

e This is significant on a one-tailed test but would not be on a two-tailed test.

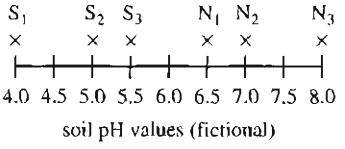
Activity 3.5, page 41

1 The null hypothesis is that there is no difference between soil samples from north-facing and south-facing slopes. On this hypothesis, the two sets of results are essentially the same and come from the same population. The students' experimental hypothesis (the alternative hypothesis) is that the two sets of results are different. They suspect that north-facing slopes will have more alkaline soils, and these will be indicated by higher pH values.

2 **a** Here are the results for set A shown on a number line:

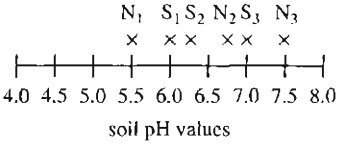


b Set B



In this case the two sets of results are totally separated.

Set C



3 Returning to the fieldwork results, U_1 and U_2 can be calculated for these:

Set A

North-facing		South-facing	
6.5	N_1	5.5	S_1
7.0	N_2	6.0	S_2
7.5	N_3	6.75	S_3
S's < N_1 = 2		N's < S_1 = 0	
S's < N_2 = 3		N's < S_2 = 0	
S's < N_3 = 3		N's < S_3 = 1	
Total $U_1 = 8$		Total $U_2 = 1$.	

Set B

S's under N's	N's under S's
S's < N_3 = 3	N's < S_1 = 0
S's < N_2 = 3	N's < S_3 = 0
S's < N_1 = 3	N's < S_2 = 0
Total $U_1 = 9$	Total $U_2 = 0$

These are the most extreme values possible for U_1 and U_2 for samples of this size.

Set C

S's under N's	N's under S's
S's < N_1 = 0	N's < S_1 = 1
S's < N_2 = 2	N's < S_2 = 1
S's < N_3 = 3	N's < S_3 = 2
Total $U_1 = 5$	Total $U_1 = 4$

4 $U_1 + U_2 = 9$ for all the sample in this activity.
$0 \le U_1 \le 9, \qquad 0 \le U_2 \le 9$
Extreme values of U_1 and U_2 indicate that the two sets of data are more separated.

Activity 3.6, page 43

1 a
FindU
Input None

$3 \to N$
$3 \to S$
$0 \to U$
repeat
int(rand + $(N/(N+S))$) $\to B$
if $B = 1$
then
$N - 1 \to N$
$U + S \to U$
else
$S - 1 \to S$
endif
until $N \le 0$
display U
Output The value of U.

b
ProbU
Input None

(2,10) $\to$ dim X
$0 \to U$
repeat
$U \to X(1, U+1)$
$0 \to X(2, U+1)$
$U + 1 \to U$
until $U = 10$
$0 \to K$
repeat
FindU
$X(2, U+1) + 1 \to X(2, U+1)$
$K + 1 \to K$
until $K = 100$
display X
Output X, the frequency table.

2 a 20

2b, c and 3b The table shows all possible sequences of S's and N's

Arrangement	U_1	U_2	Probability	
1 SSSNNN	0	9	$\frac{1}{20}$	0.05
2 SSNSNN	1	8	$\frac{1}{20}$	0.05
3 SSNNSN	2	7	$\frac{2}{20}$	0.1
4 SNSSNN	2	7		
5 SSNNNS	3	6		
6 SNSNSN	3	6	$\frac{3}{20}$	0.15
7 NSSSNN	3	6		
8 SNSNNS	4	5		
9 SNNSSN	4	5	$\frac{3}{20}$	0.15
10 NSSNSN	4	5		
11 SNNSNS	5	4		
12 NSSNNS	5	4	$\frac{3}{20}$	0.15
13 NSNSSN	5	4		
14 SNNNSS	6	3		
15 NSNSNS	6	3	$\frac{3}{20}$	0.15
16 NNSSSN	6	3		
17 NSNNSS	7	2		
18 NNSSNS	7	2	$\frac{2}{20}$	0.1
19 NNSNSS	8	1	$\frac{1}{20}$	0.05
20 NNNSSS	9	0	$\frac{1}{20}$	0.05

3 a If the null hypothesis is true, any one of the 20 possible combinations of S and N is equally likely to occur by chance. So each one has a probability of $\frac{1}{20}$.

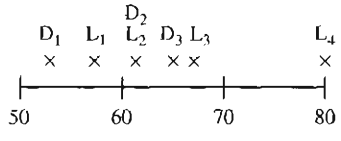
4

U	Probability
0	0.1
1	0.1
2	0.2
3	0.3
4	0.3

Activity 3.7, page 45

1 a The null hypothesis is that the percentage of brine shrimp eggs hatching under light and dark conditions is the same. The alternative hypothesis is that conditions affect the percentage of eggs hatching out so that the percentage under light conditions is higher than under dark conditions.

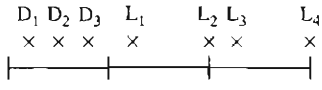
b



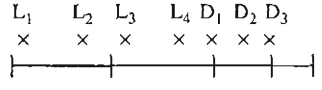
c

U_1	U_2
$D < L_1 = 1$	$L < D_1 = 0$
$D < L_2 = 1.5$	$L < D_2 = 1.5$
$D < L_3 = 3$	$L < D_3 = 2$
$D < L_4 = 3$	
$U_1 = 8.5$	$U_2 = 3.5$

d



$U_1 = 12$ $U_2 = 0$



$U_1 = 0$ $U_2 = 12$

e $\binom{7}{4} = 35$, so the probability of each is $\frac{1}{35}$.

f $\frac{2}{35} = 5.7\%$. You cannot reject the null hypothesis. Even $U = 0$ is not significant at the 5% level.

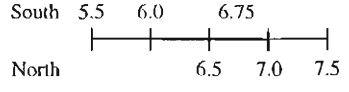
g $\binom{8}{4} = 70$, so the probability of $U = 0$ is $\frac{2}{70} = 2.9\%$, so a significant result is possible.

2 a $U_1 = 0$, $U_2 = 20$. Values of U from 0 to 10 are possible. There are $\binom{9}{5} = 126$ equally likely sequences. The probability of each one is $\frac{1}{126} = 0.8\%$, so the probability that $U = 0$ is $\frac{2}{126} = 1.6\%$.

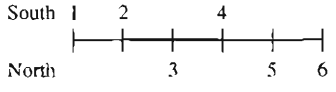
b $U_1 = 0$, $U_2 = 25$. Values of U from 0 to 12 are possible. There are $\binom{10}{5} = 252$ equally likely sequences. The probability of each one is $\frac{1}{252} = 0.4\%$, so the probability that $U = 0$ is $\frac{2}{252} = 0.8\%$.

Activity 3.8, page 46

1 a



b



c

South	$R_1 = 1 + 2 + 4 = 7$
North	$R_2 = 3 + 5 + 6 = 14$
$R_1 = 7$	$n_1 = 3$
$R_2 = 14$	$n_2 = 3$

d

$$U_1 = 7 - \left(\tfrac{1}{2} \times 3 \times 4\right)$$
$$= 7 - 6$$
$$= 1$$

$$U_2 = 14 - \left(\tfrac{1}{2} \times 3 \times 4\right)$$
$$= 14 - 6$$
$$= 8$$

These are the same results as before, but in reverse order, which does not matter.

Activity 3.9, page 48

1 a The null hypothesis is that giving swimmers a target makes no difference to their times. The alternative hypothesis is that giving a target does make a difference.
b $R_1 = 128.5$ and $R_2 = 81.5$, giving $U_1 = 73.5$, $U_2 = 26.5$ and $U = 26.5$, which is not significant at the 5% level, as the critical value is 24.

2 $R_1 = 146.5$ and $R_2 = 63.5$, giving $U_1 = 91.5$, $U_2 = 8.5$ and $U = 8.5$, which is significant at the 5% level. It is also significant at the 1% level, as the critical value is 16 so reject the null hypothesis.

3 a This is not a true experiment. The nurse collecting the results has not been able to control any part of the situation. She has merely collected data after the event. At best, this could be called a survey but her sampling methods are very dubious.
b Her samples may not be at all representative of either population (non-smoking and smoking mothers). They are a select group of women she knows – not a random sample from the whole population.
c Extraneous variables which have not been controlled:
age – are the non-smoking mothers all younger?
social class – are the smoking mothers of a different social class to the non-smokers?
diet – do the non-smokers eat more (or eat a better diet)?
standards of housing;
alcohol consumption;
racial group;
other medical factors;
marital status;
amount of time since the smokers started smoking;
number of cigarettes smoked per day;
does the husband/partner smoke?
number of children already.

All these factors can affect the birth weights of babies.

d Null hypothesis: there is no difference in birth weights of babies born to smoking and non-smoking mothers. Alternative hypothesis: there is a difference in birth weights of babies born to smoking and non-smoking mothers; with babies of smoking mothers weighing less.
e Results of experiment using a Mann-Whitney U test

$U_1 = 191$ and $U_2 = 19$

So $U = 19$, which is highly significant as the critical value at the 1% level is 47.
f The result is highly significant, and it indicates the need for further investigation. However, it doesn't prove anything.

Activity 3.11, page 50

1 **a** Independent samples
b Mann-Whitney U test
c The hypnosis treatment and the standard treatment are equally effective in treating students with examination stress.
d The new hypnosis treatment is more effective.
e $R_1 = 96.5$

$R_2 = 134.5$

$U_1 = 30.5$

$U_2 = 79.5$

$U = 30.5$

f Significant value for U for a one-tailed test at the 5% level is 31, so this result is just about significant. The counsellor should continue with the experiment and collect some more data.

2 **a** Independent samples
b Mann-Whitney U test
c Men and women are equally good at abstract reasoning questions in the test.
d Men will gain higher scores on the test.
e $R_1 = 75$

$R_2 = 61$

$U_1 = 39$

$U_2 = 25$

$U = 25$

f Significant value for U from tables is 13 so this result is not significant, so men and women are equally good at this test.

3 **a** Repeated measures taken on the same sample
b Wilcoxon signed ranks test
c The secretaries will be able to type the reports equally quickly on the electric typewriter and the word processor.
d The secretaries will not have similar typing speeds on the two machines. They may be quicker on the word processor, assuming they have all learned how to use it.
e Sum of + ranks = 4.5
Sum of – ranks = 23.5.

Eight ways of achieving $T = 4.5$ or lower probability $= \frac{8}{128} = 0.0625$. Alternatively – From Table 4 the critical value for t is 2 (or 4 for a one-tailed test) at the 5% level.
f This is not significant, even if you feel justified in using a one-tailed test.

4 **a** Repeated measures 'before and after' design on the same sample
b Wilcoxon signed ranks test
c The elderly people's self-image will be the same before and after the exercise programme.
d The self-image of the elderly subjects will improve during the exercise programme.
e Sum of + ranks = 4
Sum of – ranks = 74.

Seven ways of achieving this total or lower, with tied ranks, probability $\frac{7}{4096} = 0.0017$.
f This is significant. The self-image of the elderly people clearly did improve.

5 **a** Independent samples
b Mann-Whitney U test
c The tar content of the two brands of cigarettes is similar.
d The tar content of the new brand is lower.
e $R_1 = 31.5$

$R_2 = 73.5$

$U_1 = 3.5$

$U_2 = 45.5$

$U = 3.5$

f For a one-tailed test at the 0.5% level, the significant value for U from tables is 5, so the result is highly significant.

6 **a** Independent samples
b Mann-Whitney U test
c The right-handers and left-handers will score equally well on the test of spatial ability.
d The scores for right-handers and left-handers are different.

e $R_1 = 62$

$R_2 = 74$

$U_1 = 26$

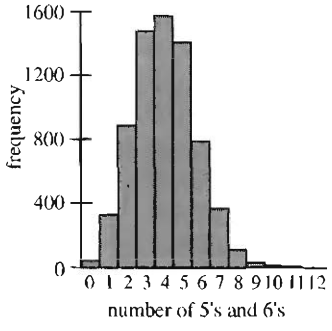
$U_2 = 38$

$U = 26$

f Significant value for U from tables is 13. For a two-tailed test at the 5% level, this result is not significant, so there is no difference between the scores for left-handed and right-handed subjects.

Activity 4.1, page 57

1 The mean number of fives and sixes is 4.04.
The proportion of fives and sixes is 0.3367.



2 Expect the proportion of throws scoring five or six to be $\frac{1}{3}$.

Null hypothesis: the number of dice out of 12 scoring five or six follows a binomial distribution with $n = 12$ and $p = \frac{1}{3}$.

Alternative hypothesis: the number of dice out of 12 scoring five or six does not follow a $B\left(12, \frac{1}{3}\right)$ distribution.

Activity 4.2, page 58

1 Assumptions:
The number of dice thrown each time is 12.
The probability of getting a five or a six is the same for each of the 12 dice, and remains the same for every throw.
The scores on the 12 dice are independent.

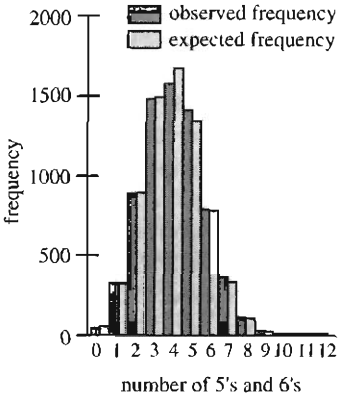
2 $n = 12$ and $p = \frac{1}{3}$

3 and 4

Number of 5s and 6s	Probability	Expected frequencies
0	0.00771	54.00
1	0.04624	323.99
2	0.12717	890.96
3	0.21195	1484.94
4	0.23845	1670.55
5	0.19076	1336.44
6	0.11127	779.59
7	0.04769	334.11
8	0.01490	104.41
9	0.00331	23.20
10	0.00050	3.48
11	0.00005	0.32
12	0.00000	0.01
Totals	1	7006.00

Activity 4.3, page 59

1



2 and 3

Number of 5s or 6s	O	E	$(O-E)$
0	45	54.00	−9.00
1	327	323.99	3.01
2	886	890.96	−4.96
3	1475	1484.94	−9.94
4	1571	1670.55	−99.55
5	1404	1336.44	67.56
6	787	779.59	7.41
7	367	334.11	32.89
8	112	104.41	7.59
9	29	23.20	5.80
10	2	3.48	−1.48
11	1	0.32	0.68
12	0	0.01	−0.01
Total	7006	7006	0.00

The total of $(O-E)$ is always zero because the totals of O and of E are the same.

4 Squaring the deviations removes all the negative signs.

5 The squared deviation for $x = 0$ is 81 on an expected frequency of only 54, while the squared deviation for $x = 3$ is about 100 on an expected frequency of nearly 900. Clearly the expected frequency for $x = 0$ differs more significantly from the corresponding observed frequency. Dividing by E gives a measure of the deviation relative to what was expected.

6 But the expected frequencies for 11 and 12 are both very close to zero, and dividing will produce large values here.

7

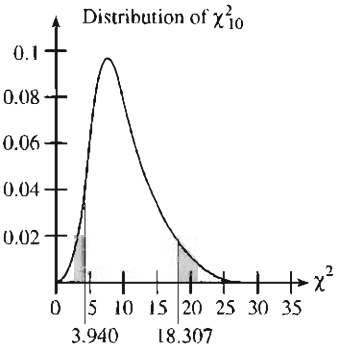
Number of 5s or 6s	O	E	$\frac{(O-E)^2}{E}$
0	45	54.00	1.50
1	327	323.99	0.03
2	886	890.96	0.03
3	1475	1484.94	0.07
4	1571	1670.55	5.93
5	1404	1336.44	3.42
6	787	779.59	0.07
7	367	334.11	3.24
8	112	104.41	0.55
9	29	23.20	1.45
10 to 12	3	3.81	0.17
Total	7006	7006	16.46

Activity 4.4, page 61

1 In this example there are eleven expected frequencies. However, as the total frequency must equal 7006, only ten of the expected frequencies can vary independently and the eleventh one must take whatever values are required to make the total equal 7006. This is called a **constraint**. To calculate ν, the number of degrees of freedom:

ν = number of classes or groups
−number of constraints

2



Distribution of χ^2_{10}

3 16.45

4 If you look at a table of χ^2 values you will find that only 5% of the results for 10 degrees of freedom are greater than 18.307. So any value of $\sum \frac{(O-E)^2}{E}$ higher than 18.307 indicates that there are discrepancies between the observed and expected frequencies. If the dice were unbiased, and the results were recorded accurately, then discrepancies of this magnitude might occur in less than five experiments out of 100. As this result has a low probability (≤ 0.05) it would be reasonable to conclude either that a biased dice was used in generating the observed frequencies, or that the recorded results contain errors. The value of X^2 generated from Weldon's data is 16.45 which is less than the critical value of 18.307 so it is not reasonable to reject the data on the strength of this test.

Activity 4.5, page 62

1 Only if the probability of scoring a five or a six is the same for each of the 12 dice.

2 Null hypothesis: the number of dice out of 12 scoring five or six follows a binomial distribution with $n = 12$ for some p.

Alternative hypothesis: the number of dice out of 12 scoring five or six does not follow a binomial distribution with $n = 12$.

3 The best estimate for p is the average proportion of times scoring five or six, which is 0.3367 from Activity 4.1 question 1.

4, 5 and 6

No. of 5s or 6s	Probability	O	E	$\frac{(O-E)^2}{E}$
0	0.00725	45	50.8	0.67
1	0.04418	327	309.5	0.99
2	0.12335	886	864.2	0.55
3	0.20871	1475	1462.2	0.11
4	0.23837	1571	1670.0	5.87
5	0.19360	1404	1356.4	1.67
6	0.11465	787	803.3	0.33
7	0.04989	367	349.5	0.88
8	0.01583	112	110.9	0.01
9	0.00357	29	25.0	0.63
10	0.00054			
11	0.00005	3	4.2	0.33
12	0.00000			
Total	1	7006	7006	12.04

7 $\nu = 9$

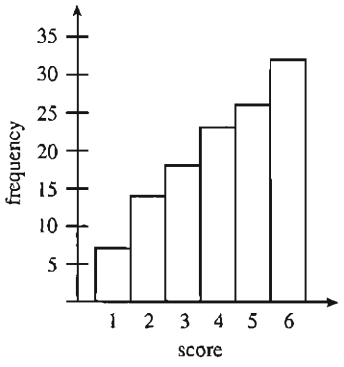
8 The critical value is $\chi^2_9(0.95) = 16.919$, from tables. The observed value here is 12.04 which is less than the critical value. You should not reject the null hypothesis.

9 The effect would be indistinguishable from all 12 dice having the same bias.

10 Suppose one or more of the 12 dice has a different bias from the others.

Activity 4.6, page 63

1 Mean is 4.19; for a theoretical uniform distribution the mean would be 3.5.



The dice appears to produce very few scores of 1 or 2, and very many of 5 or 6.

2 Expect each score about 20 times.

Null hypothesis: probability of scoring x is $\frac{1}{6}$, for each possible score; this means that the scores should be uniformly distributed with $p = \frac{1}{6}$.

Alternative hypothesis: probability of scoring x is different for each possible score.

3 $\sum \frac{(O-E)^2}{E} = 19.9$

4 *DiceFreq*

Input M

$120 \rightarrow M$
$(2,6) \rightarrow \dim X$
for $J = 1$ **to** 6
  $J \rightarrow X(1,J)$
  $0 \rightarrow X(2,J)$
next
for $J = 1$ **to** M
  $\text{int}(1 + 6\text{rand}) \rightarrow D$
  $1 + X(2,D) \rightarrow X(2,D)$
next
Output X

6 *ChiSquare*
Input N

for $I = 1$ **to** N
$0 \rightarrow C$
DiceFreq
for $K = 1$ **to** 6
$C + (X(2, K) - 20)^2 \div 20 \rightarrow C$
next
display C
pause
next
Output A list of numbers C

In the algorithm above, N is the number of times *DiceFreq* is called. The 'pause' command has been used to enable people using a graphics calculator to record the value of C each time.

The distribution is asymmetric, clustered towards the lower values, with a long tail to the right.

7 If you are testing at the 5% level, then, from the test, you would expect a fair dice to appear as biased about one time in twenty. This is the probability of producing a significant result when the null hypothesis is really true.

Activity 4.7, page 64

1 Null hypothesis: the preferences expressed for each brand are roughly equal; that is to say uniform distribution $p = \frac{1}{3}$.
Alternative hypothesis: there are not equal preferences for the three brands.
$\nu = 2$.
Expected frequencies are all 50.

$$\sum \frac{(O-E)^2}{E} = 43.0$$

This result is greater than 5.99, the value given in tables for χ^2. So the brands of cola are not chosen with equal preference.

2 Null hypothesis: people think fat people and thin people are equally jolly.
Alternative hypothesis: people do not think that fat people and thin people are equally jolly.
$\nu = 1$
Expected frequencies are both 40.

$X^2 = 2.45$
This result is much less than $\chi_1^2(5\%) = 3.84$ from tables, so the distribution is not significantly different from the predicted.

3 Null hypothesis: each of the four wrappers is equally popular with customers.
Alternative hypothesis: one or more of the wrappers is preferable to the others.
$\nu = 3$
The expected frequencies are 101.25 each, giving a total frequency of 405 which matches what was observed.

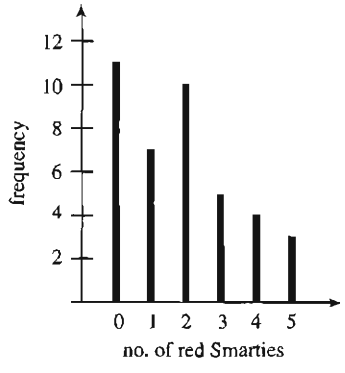
$X^2 = 12.31$
This result is greater than 7.81, the value from tables of $\chi_3^2(5\%)$. So the observed distribution is significantly different from the uniform distribution expected under the null hypothesis.

4 Null hypothesis: the four types occur in the ratio 9:3:3:1.
Alternative hypothesis: the four types do not occur in these proportions.
$\nu = 3$
Expected frequencies are 90, 30, 30, 10.

$X^2 = 4.51$
This result is less than 7.81 so the distribution is not significantly different from the predicted model.

Activity 4.8, page 66

1 Mean: 1.825; variance 2.444.



2 Theoretically p should equal 0.125 if there are equal numbers of all the eight colours. n is 16.

Null hypothesis: the distribution of red Smarties is distributed binomially with $n = 16$ and $p = \frac{1}{8}$.
Alternative hypothesis: the distribution of red Smarties is not distributed binomially with $n = 16$ and $p = \frac{1}{8}$.

3 Binomial probabilities:

x	p(x)
0	0.118
1	0.270
2	0.289
3	0.193
4	0.089
5	0.031
6 or more	0.010

4 Expected frequencies for values of 0, 1, 2, 3, 4, 5, and 6 or more: 4.7, 10.8, 11.6, 7.7, 3.6, 1.2, 0.4.

5 $\sum \frac{(O-E)^2}{E} = 11.573$

6 Five classes; with one constraint, that the total frequency = 40; $\nu = 4$.

7 9.49. Yes, it is significant.

8 Red Smarties do not fit a binomial distribution $p = \frac{1}{8}$ when fun-size boxes are investigated in this sample. Note that the mean should be:

$\bar{x} = n \times p$

$\bar{x} = 16 \times \frac{1}{8}$

$\bar{x} = 2$

The mean in this study is 1.825. You could use this to calculate a lower value of p and test again for a different binomial distribution using the method outlined in Activity 4.7. Or it may be that the assumptions needed to model with a binomial distribution are not valid.

Activity 4.9, page 67

1 $\bar{x} = 2.48$.

2 $p = 0.248$.

3 Null hypothesis: the number of drawing pins landing point-up in a group of ten is distributed as a binomial distribution with $n = 10$ and $p = 0.248$. Alternative hypothesis: the number of drawing pins landing point-up is not distributed as a binomial distribution with $n = 10$ and $p = 0.248$.

4 and **5**

x	p(x)	Expected frequency
0	0.058	11.6
1	0.191	38.2
2	0.283	56.6
3	0.249	49.8
4	0.144	28.8
5	0.057	11.4
6 or more	0.018	3.6

6 **a** One group has $E \leq 5$.
b Yes
c Six classes.
d $\nu = 4$.

7

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
11	11.6	−0.6	0.36	0.031
39	38.2	0.8	0.64	0.017
47	56.6	−9.6	92.16	1.628
53	49.8	3.2	10.24	0.206
46	28.8	17.2	295.84	10.272
4	15.0	−11.0	121.00	8.066
			Total =	20.22

8 9.49

9 Yes

10 The results do not fit the binomial distribution B (10, 0.248). This implies that one or more of the assumptions for the binomial distribution is not true here.

Activity 4.10, page 68

	Mean number of pins landing point-up out of ten	Probability point-up
Student A	7.04	0.704
Student B	7.06	0.706
Student C	7.04	0.704
Student D	6.97	0.697
Student E	6.60	0.660
Student F	6.66	0.666
Student G	5.10	0.510
All students	6.63	0.663
Exclude G	6.86	0.686

The mean number of pins point-up appears to be much lower for student G than for the other students.

2 Null hypothesis: the data from student G follows the same distribution as the data from the other students. Alternative hypothesis: data from student G follows a different distribution from the other students.

3 a Chi-squared test including student G

x	O	E	$\frac{(O-E)^2}{E}$
0–3	18	8.08	12.17
4	24	22.92	0.05
5	58	54.08	0.28
6	60	88.60	9.23
7	93	99.55	0.43
8	89	73.39	3.32
9	36	32.07	0.48
10	7	6.30	0.08
Total	385	385.00	26.04

Test using chi-squared with six degrees of freedom, 26.04 is highly significant (even at the 0.1% level of significance).

3 b Chi-squared test excluding student G

x	O	E	$\frac{(O-E)^2}{E}$
0-3	4	4.68	0.10
4	15	15.00	0.00
5	52	39.26	4.13
6	53	71.37	4.73
7	85	88.96	0.18
8	87	72.77	2.78
9	33	35.28	0.15
10	6	7.69	0.37
Total	335	335.00	12.44

The critical value of chi-squared with six degrees of freedom at the 5% level is 12.59, so the value 12.44 is not significant at the 5% level. But 12.44 is quite close to the critical value, so be aware that the fit may not be good.

4 Chi-squared test of student G results against the distribution which was fitted to the results of all students excluding student G.

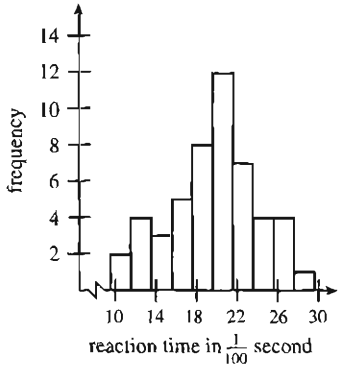
x	O	E	$\frac{(O-E)^2}{E}$
0-4	23	2.94	137.09
5	6	5.86	0.00
6	7	10.65	1.25
7	8	13.28	2.10
8	2	10.86	7.23
9	3	5.27	0.97
10	1	1.15	0.02
Total	50	50.00	148.66

The value of 148.66 is highly significant on the chi-squared distribution with five degrees of freedom.

It would appear that the data collected by student G were collected under different conditions from the data collected by the other students. Student G has too many occasions when the number of pins pointing up is less than five. One possible explanation for this might be that student G threw the ten pins on to a different kind of surface, one which allowed the pins to bounce more before they finally landed. If the pins bounce more, then they might be expected to be more likely to land with the heaviest end downwards (that is point-up). This might occur if the other students dropped their pins onto a pad of paper and student G dropped them onto a bare desk. This explanation of course raises questions about the validity of the binomial model for the other data as well, since you do not know how well the experimental conditions were controlled by the other students.

Activity 4.11, page 69

1



2 a $\bar{x} = 19.62$

b $\sigma = 4.33$

3

Upper limit	Z	Area to left	Probability	Expected frequency
9.5	−2.34	0.00964	0.00964	0.5
11.5	−1.88	0.03005	0.02041	1.0
13.5	−1.41	0.07927	0.04922	2.5
15.5	−0.95	0.17106	0.09179	4.6
17.5	−0.49	0.31207	0.14101	7.1
19.5	−0.03	0.48803	0.17596	8.8
21.5	0.43	0.66640	0.17837	8.9
23.5	0.90	0.81594	0.14954	7.5
25.5	1.36	0.91309	0.09715	4.9
27.5	1.82	0.96562	0.05253	2.6
29.5	2.28	0.98870	0.02308	1.2
∞	∞	1.00000	0.01130	0.6

4 Null hypothesis: the reaction times follow a normal distribution with mean 19.62 and standard deviation 4.33 hundredths of a second.

Alternative hypothesis: the reaction times do not fit a normal distribution with mean 19.62 and standard deviation 4.33 hundredths of a second.

5

Upper limit	E		O	$\frac{(O-E)^2}{E}$
9.5	0.5			
11.5	1.0	4	6	1
13.5	2.5			
15.5	4.6		3	0.556
17.5	7.1		5	0.621
19.5	8.8		8	0.073
21.5	8.9		12	1.080
23.5	7.5		7	0.033
25.5	4.9		4	0.165
27.5	2.6			
29.5	1.2	4.4	5	0.082
∞	0.6			
			Total	3.610

If the end four cells are merged, the results become

Range	E	O	$\frac{(O-E)^2}{E}$
<15.5	8.6	9	0.019
≥23.5	9.3	9	0.010
		Total	1.836

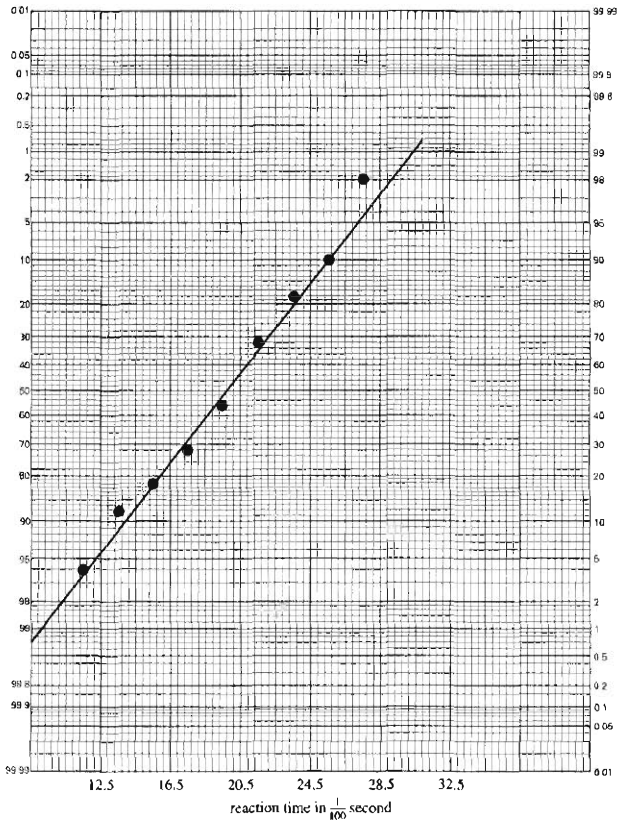
6 $\chi^2_5 = 11.07$ and $\chi^2_3 = 7.815$ at the 5% level. Both methods of calculating X^2 in question 5 show that the distribution of reaction times does not differ significantly from a normal distribution.

Activity 4.12, page 72

1 and 2

Upper limit	f	Cumulative frequency	Cumulative relative frequency	%
11.5	2	2	0.04	4
13.5	4	6	0.12	12
15.5	3	9	0.18	18
17.5	5	14	0.28	28
19.5	8	22	0.44	44
21.5	12	34	0.68	68
23.5	7	41	0.82	82
25.5	4	45	0.90	90
27.5	4	49	0.98	98
29.5	1	50	1.00	100

3 and 4



5 From the graph drawn here, the mean reaction time is estimated at about 19.2 hundredths of a second.

6 From the graph, the corresponding value is found to be about 23.4 hundredths of a second, although estimates based on calculations directly from the data are likely to be better.

Standard deviation= 23.4 – 19.2 = 4.2

Activity 4.13, page 72

1

Upper limit	f	Cumulative frequency	Cumulative relative frequency	%
53.5	1	1	0.0095	0.95
57.5	1	2	0.0190	1.90
61.5	5	7	0.0667	6.67
65.5	12	19	0.1810	18.10
69.5	23	42	0.4000	40.00
73.5	25	67	0.6381	63.81
77.5	20	87	0.8286	82.86
81.5	13	100	0.9524	95.24
85.5	4	104	0.9905	99.05
89.5	1	105	1.0000	100.00

Pulse rates:
mean estimated from graph is 71 beats per minute; the standard deviation estimated from graph is
77.5 – 71 = 6.5

2 The points will not lie on a straight line.

Activity 4.15, page 76

1 Mean from graph is 65 mm
The standard deviation from the graph is
72 – 65 = 7 mm.

2 From calculator: mean = 64.4; $\sigma = 6.63$

Activity 4.16, page 76

1 Null hypothesis: vegetarians and non-vegetarians are equally likely to smoke.

Alternative hypothesis: vegetarians are less likely than non-vegetarians to smoke.

2

Observed frequencies	Vegetarian	Non-vegetarian	Totals
Smokers	5	21	26
Non-smokers	27	42	69
Totals	32	63	95

3

Observed frequencies	Vegetarian	Non-vegetarian	Totals
Smokers	8.8	17.2	26
Non-smokers	23.2	45.8	69
Totals	32	63	95

4

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
5	8.8	–3.8	14.44	1.64
27	23.2	3.8	14.44	0.84
21	17.2	3.8	14.44	0.62
42	45.8	–3.8	14.44	0.32
Total				3.42

5 It does not matter which one of the expected frequencies you choose, but once one expected frequency is entered into a two-by-two table, all the others are determined by the marginal totals. So for a two-by-two table, there is only one degree of freedom.

6 a Significant value of χ_1^2 is 3.84.
b No
c Smoking is equally prevalent among vegetarians and non-vegetarians.

7 a 2
b In each column, the last figure will be determined by the column total, so that in each column there are only $m-1$ free choices, rather than m. Likewise, in each row, the last figure will be determined by the row total. In each row there are only $n-1$ free choices rather than n. The number of degrees of freedom for a two-way contingency table with m rows and n columns can be found by using $\nu=(m-1)(n-1)$.

Activity 4.17, page 79

1 Null hypothesis: people are equally likely to wear their watches on their writing hands as their non-writing hands.
Alternative hypothesis: people are more likely to wear their watches on the hands they do not use for writing.

Expected frequencies 16.4, 33.6; 6.6, 13.4.

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
11	16.4	−5.4	29.16	1.78
12	6.6	5.4	29.16	4.42
39	33.6	5.4	29.16	0.87
8	13.4	−5.4	29.16	2.18
Total				9.25

$\nu=1$

The result is significant.

Conclude that people are more likely to wear their watch on their non-writing hands.

2 Null hypothesis: children who have been vaccinated are just as likely to catch measles as children who have not been vaccinated.
Alternative hypothesis: children who have been vaccinated are less likely to catch measles.

Expected frequencies: 22, 78; 22, 78.

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
10	22	−12	144	6.55
34	22	12	144	6.55
90	78	12	144	1.85
66	78	−12	144	1.85
Total				16.80

$\nu=1$

The result is significant.

Conclude that the children who have been vaccinated are less likely to catch measles.

3 Null hypothesis: staff in different categories all experience the same degree of job satisfaction.
Alternative hypothesis: staff in unqualified jobs experience less job satisfaction.

Expected frequencies: 43.1, 36.9; 48.5, 41.5; 48.5, 41.5

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
73	43.1	29.9	894.01	20.74
45	48.5	−3.5	12.25	0.25
22	48.5	−26.5	702.25	14.48
7	36.9	−29.9	894.01	24.23
45	41.5	3.5	12.25	0.30
68	41.5	26.5	702.25	16.80

Total for $\sum\frac{(O-E)^2}{E}=76.92$.

$\nu=2$. Significant value $\chi^2=5.99$

The result is significant.

Conclude that all staff do not experience equal levels of job satisfaction.

4 Null hypothesis: hair colour and eye colour are not related.
Alternative hypothesis: there is a relationship between hair colour and eye colour.

Expected frequencies:

Hair colour	Eye colour		
	Blue	Brown	Black
Black	13.9	21	7.1
Brown	29.2	44	14.8
Red	3.6	5.5	1.9
Blonde	10.3	15.5	5.2

Combined red and blonde hair (if you think that is valid).

Hair colour	Eye colour		
	Blue	Brown	Black
Black	13.9	21	7.1
Brown	29.2	44	14.8
Blonde or Red	13.9	21	7.1

O	E	$O-E$	$\frac{(O-E)^2}{E}$
12	13.9	−1.9	0.26
21	29.2	−8.2	2.30
24	13.9	10.1	7.34
21	21	0	0
54	44	10	2.27
11	21	−10	4.76
9	7.1	1.9	0.51
13	14.8	−1.8	0.22
7	7.1	−0.1	0.00

$$\sum \frac{(O-E)^2}{E} = 17.66$$

$\nu = 4$ for reduced table
Significant value of $\chi^4_2 = 9.49$

This result is significant.

Conclude that hair colour and eye colour are related.

Activity 5.2, page 84

4 Discrete
The variable is the result of counting the number of people (or cars) arriving in a fixed time interval.

Activity 5.4, page 86

2 You should find that the values of the binomial and Poisson probabilities are very close to one another.

5 The model assumes that births are equally spread throughout the year, so that the probability that a randomly chosen person has a particular birthday is equally likely for all days in the year. In fact, births are subject to some seasonal variation, so this assumption is not really valid.

Furthermore, your sample is drawn from registers of sixth form students, and this may involve some selection by date of birth. Young people born in August have been in full-time education almost 12 months less than those born in the previous September, yet they are all in the same school year. You might suppose that young people born in September, who have been in school longer, are either more or less likely to stay on for the sixth form than those in their year who are younger.

Activity 5.6, page 89

1 **a** $e^{-5} = 0.0067$
b $5e^{-5} = 0.0337$
c 0.735

2 3600 particles in 2 hours is equivalent to 5 in 10 seconds.

Number of particles in 10 seconds	Poisson probability
0	0.0067
1	0.0337
2	0.0842
3	0.1404
4	0.1755
5	0.1755
6	0.1462
7	0.1044
8	0.0653
9	0.0363
10	0.0181
11	0.0082
12	0.0034

3 **a** $e^{-2}\left(\frac{2^4}{4!}\right) = 0.0902$
b 0.0527

To use the Poisson distribution here you need to assume:
faults occur at random
faults only ever occur one at a time
faults occur independently
faults occur at a uniform rate.

In fact, the assumption of independence is questionable, since if a fault occurs which puts the computer out of action for a period of time, then no further fault can occur until this first fault is rectified. It is also likely that a small fault may trigger the occurrence of further faults.

4 **a** 0.2240
b 0.3258

The Poisson model is very unlikely to be appropriate here since such scholarships are likely to occur in bunches. Good students in a year group might be expected to encourage each other to apply for industrial scholarships, and to succeed.

5 The stock should be at least 10 bulbs.

Activity 5.7, page 89

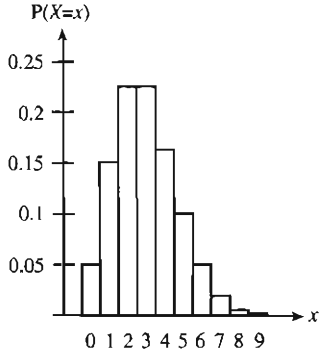
1 Null hypothesis: the new working practices have made no difference; the mean number of accidents per month is still three.

Alternative hypothesis: the new regulations have brought about a decrease in the mean number of accidents per month.

$H_0: \mu = 3$ $H_1: \mu < 3$

2

x	$P(X = x)$
0	0.0498
1	0.1494
2	0.224
3	0.224
4	0.168
5	0.1008
6	0.0504
7	0.0216
8	0.0081
9	0.0027



3 $P(X \leq 1) = 0.0498 + 0.1494 = 0.1992$

This is not significant. On the basis of this one month's results, it is not possible to say whether the number of accidents has been reduced. If there had been no accidents that month, that would have been a significant result.

Activity 5.8, page 90

1 Null hypothesis: the new arrangements have made no difference; the mean number of accidents in six months is still six.

Alternative hypothesis: the new arrangements have brought about a decrease in the mean number of accidents in six months.

$H_0: \mu = 6$ $H_1: \mu < 6$

If H_0 is true, then the probability of two or fewer accidents in a six-month period is 0.0620.

Conclude that two accidents in six months is not conclusive evidence, but there is some indication that there may have been an improvement. A continued trial over a longer period, say one or two years, might be more conclusive.

2 Null hypothesis: the frost has made no difference; the mean number of calls per day is still one.

Alternative hypothesis: the frost has brought about an increase in the mean number of calls per day.

$H_0: \mu = 1$ $H_1: \mu > 1$

If H_0 is true, then the probability of four or more calls in one day is 0.0190.

Conclude that there is only about a two per cent chance of getting a result as extreme as this if H_0 is true. Therefore this result is significant evidence at the 5% level that the rate of calls has increased above one per day during the frost.

3 Null hypothesis: the new supplier has made no difference; the mean number of bad tomatoes per day is still two.

Alternative hypothesis: the new supplier has brought about an increase in the mean number of bad tomatoes per day.

$H_0: \mu = 2$ $H_1: \mu > 2$

If H_0 is true then the probability of six or more bad tomatoes in one day is 0.0166.

Conclude that there is less than a two per cent chance of getting a result as extreme as this if H_0 is true. Therefore this result is significant evidence at the 5% level that the number of bad tomatoes supplied by the new supplier is greater than two per day.

4 Two letters per week is equivalent to $\frac{2}{6}$ letters per day, assuming that letters can arrive six days per week.

Null hypothesis: the mean number of letters per day is still $\frac{2}{6}$.

Alternative hypothesis: the mean number of letters per day has increased.

$H_0: \mu = \frac{2}{6}$ $\quad$ $H_1: \mu > \frac{2}{6}$

If H_0 is true, then the probability of five or more letters in one day is 0.000 026 01.

Conclude that there is negligible chance of getting a result as extreme as this if H_0 is true. Therefore this result is highly significant evidence that the number of letters arriving on this day is not an observation from a Poisson distribution with mean $\frac{2}{6}$. It is likely that this is a special day for some reason. Perhaps it's the student's birthday!

Activity 5.9, page 90

1 $P(X = x) = \left(\frac{\mu}{x}\right) P(X = x - 1)$

2 and 3

x	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 10$
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0000
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0005
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0023
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0076
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.0189
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.0378
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.0631
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.0901
8	0.0000	0.0009	0.0081	0.0298	0.0653	0.1126
9	0.0000	0.0002	0.0027	0.0132	0.0363	0.1251
10	0.0000	0.0000	0.0008	0.0053	0.0181	0.1251

Activity 5.10, page 91

1 See the text following Activity 5.5.

2 The mean

3 Skewed, with more small values than large values.

4 Mean = 0.917, variance = 1.043

5

x	$P(X = x)$
0	0.3998
1	0.3665
2	0.1680
3	0.0513
4	0.0118
5	0.0022

6 23.9, 22.0, 10.1, 3.1, 0.7, 0.1

7 H_0: customers arrive independently and randomly, at a constant average rate.

H_1: not H_0.

8 The total frequency must equal 60.
The mean has to be estimated using the value of $\sum fx$ from the data.

9

x	O	E	$\frac{(O-E)^2}{E}$
0	25	23.9	0.0506
1	20	22.0	0.1818
2	13	10.1	0.8327
3 or more	2	4.0	1
Total	60	60	2.0651

10 $\chi^2_2(5\%) = 5.991$

11 The result is not significant. The Poisson model fits the data well, but the sample size of 60 is quite small.

Activity 5.11, page 92

1

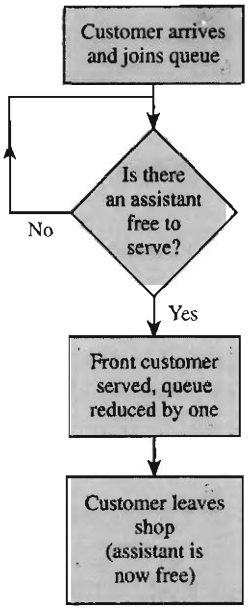
Number of calls per half hour	frequency	$P(X \geq x)$
0	1	1
1	2	$\frac{16}{17} = 0.941$
2	6	$\frac{14}{17} = 0.824$
3	6	$\frac{8}{17} = 0.471$
4	2	$\frac{2}{17} = 0.118$

2 From the graph the mean may be estimated as 2.5.

3 Calculated mean is 2.35.

Activity 6.1, page 97

1



In this example, the important processes are the times when customers arrive at the shop and the times taken to serve them; these will together determine their departure times.

2 It may be that customers' arrivals in each one minute interval (or 30 second interval) can be modelled by a Poisson distribution. Time taken to serve customers will be more difficult, but a normal distribution may be suitable.

3 If you use a Poisson distribution for the number of arrivals in a minute, then you will need to estimate the mean from the observations. When collecting data on times taken to serve customers, you may find it helpful to look at the till roll, as it often shows the time of day.

4 Data will need to be collected to investigate the times of arrival of potential customers at the shop, and the times taken to serve customers. Data about times of arrival could be collected by observing the number of potential customers arriving in each one minute interval (or 30 second interval) for an hour. As well as people actually entering the shop, you should also count those who go away because the queue is too long. Service time may be more difficult to measure and record, but you might be able to get useful data from the till roll as it often shows the time of day for each transaction.

Activity 6.2, page 98

1 Simulation of observations from a Poisson distribution $\mu = 2$

Random number (y)	Value of x
$0.0000 \le y < 0.1353$	0
$0.1353 \le y < 0.4060$	1
$0.4060 \le y < 0.6767$	2
$0.6767 \le y < 0.8571$	3
$0.8571 \le y < 0.9473$	4
$0.9473 \le y < 0.9834$	5
$0.9834 \le y < 0.9955$	6
$0.9955 \le y < 0.9989$	7
$0.9989 \le y < 0.9998$	8
$0.9999 = y$	9

2 Simulation of observations from a binomial distribution $B\left(10, \frac{1}{4}\right)$

Random number (y)	Value of x
$0.0000 \le y < 0.0563$	0
$0.0563 \le y < 0.2440$	1
$0.2440 \le y < 0.5256$	2
$0.5256 \le y < 0.7759$	3
$0.7759 \le y < 0.9219$	4
$0.9219 \le y < 0.9803$	5
$0.9803 \le y < 0.9965$	6
$0.9965 \le y < 0.9996$	7
$0.9996 \le y$	8

Activity 6.3, page 99

1 $\phi(Z) = 0.4286$

This corresponds to a negative value of Z, so subtract from 1:

$1 - 0.4286 = 0.5714$

Hence $Z = 0.18$ from tables.

Required value of $Z = -0.18$ is the observation from $N(0,1)$.

$x = 2\,\text{min}\,30\text{s} - 0.18 \times 25\text{s}$

$x = 2\,\text{min}\,25.5\text{s}$

2 and 3 Students' own simulation

Practice exercises

Chapter 3, page 101

1

Student	A	B	C	D	E
Difference	1	−2	2	3	3
Rank	2.5	5.5	5.5	7.5	7.5
Sign	+	−	+	+	+

Student	F	G	H	I	J
Difference	−1	4	0	1	−1
Rank	2.5	9	–	2.5	2.5
Sign	−	+	−	+	−

Total positive ranks = 34.5
Total negative ranks = 10.5
Therefore $T = 10.5$ and $n = 9$.
Null hypothesis: scores after are the same as scores before.
Alternative hypothesis: scores after are better than scores before.
Using a one-tailed test, because you are only looking for improvement, the critical value of T with $n = 9$ at the 5% level is 8.
The observed value is not less than 8, so there is not a significant improvement at the 5% level.

2

	Ranks						
A	2	3	7	8	11	12	13
B	1	4	5	6	9	10	

$R_1 = 56 \quad U_1 = 56 - \frac{1}{2} \times 7 \times 8 = 28$

$R_2 = 35 \quad U_2 = 35 - \frac{1}{2} \times 6 \times 7 = 14$

$\mu = 14$

H_0: there is no difference.

H_1: median pressure for B ≠ median for A. You need a test with $n = 7$ and $n = 6$. The critical value is 35. This is the upper. Lower tail value is $7 \times 6 - 35 = 7$. Since observed value of U is not less than 7 you conclude that there is no significant difference between the barometric pressures at the two sites.

3

Patient	A	B	C	D	E	F	G	H	I	J
Difference	−5	−8	−6	−7	+1	−2	+3	0	+1	−2
Rank	6	9	7	8	1.5	3.5	5	×	1.5	3.5
Sign	−	−	−	−	+	−	+	×	+	−

Total positive ranks 8
Total negative ranks 37
Therefore $T = 8$ and $n = 9$

Null hypothesis: exercise regime will have no effect on patients' fitness.

Alternative hypothesis: exercises will improve patients' fitness.

Using a one-tailed test the critical value for T is 8.

Since the observed value of T is 8 further investigation is required before any judgement can be made.

Chapter 4, page 102

1 a

	Observed	Expected	$\frac{(O-E)^2}{E}$
White front	18	24	1.5
No white front	78	72	0.5
Total	96	96	2.0

$\chi^2(0.95) = 3.841$. Hence no significant difference between observed and expected.

b

Day	Observed	Expected	$\frac{(O-E)^2}{E}$
Monday	190	241	10.79
Tuesday	278	241	5.68
Wednesday	228	241	0.70
Thursday	252	241	0.50
Friday	225	241	1.06
Saturday	273	241	4.25
Total	1446	1446	22.98

$\nu = 5$; $\chi^2_5(0.95) = 11.070$. Hence there is significant evidence that the number of books borrowed is greater on some days than others.

2

	O	P	E	$\frac{(O-E)^2}{E}$
Even and large	375	2/6	400	1.56
Even and small	183	1/6	200	1.45
Odd and large	212	1/6	200	0.72
Odd and small	430	2/6	400	2.25
Totals	1200		1200	5.98

$\chi^2_3(0.95) = 7.815$. This is not conclusive evidence of bias.

3 **a** Mean $\frac{\sum fx}{\sum f} = 3.974$. Hence estimate for p is 0.4968. Expected frequencies using $B(8, 0.4968)$ for $x = 0$ and 1 are small, so you need to combine them with $x = 2$. Similarly for $x = 7$ and 8. This leaves five cells, and there are two constraints to fit the binomial distribution, so $\nu = 3$. $\sum \frac{(O-E)^2}{E} = 5.942$. This is less than the value of χ^2_3 (5%) which is 7.815. Conclude that the binomial model is a good fit.

b Mean = 3.974; standard deviation = 1.776. You need to combine cells $x = 0$ and $x = 1$. Similarly cells $x = 7$ and $x = 8$. This leaves seven cells. There are three constraints to fit the normal distribution. Hence $\nu = 4$.

χ^2_4 (5%) = 9.488. $\sum \frac{(O-E)^2}{E} = 1.383$ which is less than the critical value. Conclude that the normal distribution is a good model.

Chapter 5, page 103

1 Let the random variable X be the number of deaths at work in 2 years. The average number of deaths in 2 years is $2 \times 0.7 = 1.4$.

Hence $X \sim \text{Po}(1.4)$. The number of deaths recorded in 2 years is 6.

$P(X < 6) = 0.9968$

$P(X \geq 6) = 0.0032$

This result is extremely unlikely if the death rate has not changed, and there is therefore significant evidence of an increase.

2 **a** From 1875 to 1894 is 20 years, and there were 14 corps, so there were $20 \times 14 = 280$ observations of the number of men killed.

b The mean number of men killed per year was 0.7 so the best estimate for μ is 0.7.

No. of men killed	$P(X = 2)$	E		O		$\frac{(O-E)^2}{E}$
0	0.4966	139.044		144		0.1766
1	0.3476	97.331		91		0.4118
2	0.1217	34.066		32		0.1253
3	0.0284	7.9	9.559	11	13	1.2387
4	0.0050	1.4		2		
5 or more	0.0007	0.2		0		
Total						1.9524

There are 3 degrees of freedom, so this result is not significant. The Poisson distribution is a good model for the data.

Appendix A The HEIGHT dataset

Extract from *Exploring Data – an introduction to data analysis for social scientists* by Catherine Marsh, Polity Press, 1988.

Height

This is a random sample drawn from the OPCS study of the heights and weights of the adult population of Great Britain in 1980 (Knight 1984). It represents a sample of two hundred married men and their wives. Nine pieces of information are recorded about each couple, as detailed below. (–1 indicates that information is missing.)

Key to columns in the dataset

1 Husband's age
2 Registrar-General's social class of the husband:
   1 class I: higher professional and managerial
   2 class II: lower professional and managerial
   3 class IIIN: routine non-manual
   4 class IIIM: skilled manual
   5 class IV: semi-skilled manual
   6 class V: unskilled manual
3 Height of the husband in millimetres
4 Wife's age
5 Registrar-General's social class of the wife (as column 2)
6 Height of wife in millimetres
7 The type of tenancy which they have:
   1 live rent free
   2 own property on mortgage
   3 own property outright
   4 council house tenant
   5 private rented accommodation
8 The age of the husband at the time of this marriage
9 Whether the husband smokes or not:
   0 did not smoke today
   1 smoked today

1	2	3	4	5	6	7	8	9
49	5	1809	43	5	1590	4	25	1
25	−1	1841	28	5	1560	5	19	1
40	6	1659	30	5	1620	4	38	1
52	6	1779	57	−1	1540	5	26	1
58	−1	1616	52	−1	1420	4	30	1
32	4	1695	27	5	1660	4	23	1
43	4	1730	52	−1	1610	4	33	1
42	3	1753	−1	−1	1635	4	30	1
47	4	1740	43	5	1580	4	24	1
31	4	1685	23	−1	1610	4	26	1
26	2	1735	25	−1	1590	2	23	0
40	1	1713	39	−1	1610	2	23	1
35	2	1736	32	2	1700	2	31	0
45	1	1715	−1	−1	1522	2	41	1
35	2	1799	35	3	1680	2	19	0
35	3	1785	33	6	1680	2	24	0
47	−1	1758	43	5	1630	5	24	1
38	−1	1729	35	5	1570	4	27	1
33	3	1720	32	3	1720	2	28	0
32	3	1810	30	5	1740	2	22	1
38	2	1725	40	−1	1600	2	31	0
45	5	1764	−1	−1	1689	4	24	1
29	4	1683	29	−1	1600	2	25	1
59	2	1585	55	2	1550	3	23	0
26	5	1684	25	5	1540	4	18	0
50	4	1674	45	5	1640	2	25	1
49	1	1724	44	5	1640	2	27	0
42	5	1630	40	−1	1630	2	28	1
33	4	1855	31	−1	1560	2	22	1
31	5	1796	−1	−1	1652	4	25	1
27	4	1700	25	−1	1580	4	21	0
57	3	1765	51	2	1570	2	32	0
34	2	1700	31	3	1590	2	28	1
28	2	1721	25	3	1650	2	23	0
46	3	1823	−1	−1	1591	2	−1	0
37	4	1829	35	3	1670	2	22	0
56	3	1710	55	3	1600	2	44	0
27	2	1745	23	3	1610	2	25	0
36	3	1698	35	2	1610	2	22	1
31	4	1853	28	3	1670	2	20	1
57	4	1610	52	6	1510	2	25	0
55	3	1680	53	5	1520	2	21	0
47	4	1809	43	5	1620	4	25	1
64	3	1580	61	−1	1530	4	21	1
60	4	1600	−1	−1	1451	4	26	1
31	4	1585	23	−1	1570	4	28	0
35	2	1705	35	−1	1580	2	25	0
36	1	1675	35	−1	1590	2	22	0
40	2	1735	39	−1	1670	2	23	0
30	4	1686	24	5	1630	2	27	0
32	4	1768	29	6	1510	4	21	1
27	3	1721	−1	−1	1560	2	26	0
20	4	1754	21	3	1660	2	19	0
45	4	1739	39	3	1610	2	25	1
59	4	1699	52	3	1440	2	27	0
43	3	1825	52	3	1570	2	25	0
29	5	1740	26	2	1670	2	24	0
48	4	1704	−1	−1	1635	2	27	1
39	4	1719	−1	−1	1670	2	25	1
47	4	1731	48	2	1730	2	21	0
54	4	1679	53	3	1560	3	−1	1
43	1	1755	42	3	1590	2	20	1
54	4	1713	50	−1	1600	4	23	1
61	4	1723	64	−1	1490	2	26	0
27	4	1783	26	3	1660	4	20	0
51	5	1585	−1	−1	1504	5	50	0
27	4	1749	32	−1	1580	5	24	0
32	4	1710	31	−1	1500	2	31	1
54	4	1724	53	5	1640	4	20	0
37	4	1620	39	−1	1650	4	21	1
55	4	1764	45	6	1620	4	29	0
36	2	1791	33	−1	1550	2	30	1
32	2	1795	32	1	1640	2	25	−1
57	5	1738	55	5	1560	4	24	0
51	4	1639	−1	−1	1552	2	25	1
62	2	1734	−1	−1	1600	3	33	0
57	2	1695	−1	−1	1545	2	22	0
51	2	1666	52	2	1570	5	24	1
50	5	1745	50	2	1550	1	22	0
32	3	1775	32	2	1600	2	20	1
54	5	1669	54	6	1660	3	20	0
34	4	1700	32	3	1640	2	22	0
45	−1	1804	41	3	1670	2	27	0
64	2	1700	61	−1	1560	3	24	1
55	4	1664	43	3	1760	3	31	0
28	5	1750	41	−1	1550	4	21	1
27	5	1753	28	−1	1640	4	23	1
55	4	1788	51	5	1600	2	26	0
27	4	1765	−1	−1	1571	4	−1	1
41	6	1680	41	5	1550	4	22	1
44	2	1715	41	3	1570	3	24	0
22	5	1755	21	3	1590	2	21	0
30	3	1764	28	3	1650	5	29	1
53	2	1793	47	3	1690	2	31	0
42	3	1731	37	3	1580	2	23	1
31	4	1713	28	3	1590	2	28	1
36	4	1725	35	5	1510	4	26	1
56	3	1828	55	−1	1600	3	30	0
46	4	1735	45	2	1660	2	22	1
34	2	1760	34	3	1700	2	23	0

1	2	3	4	5	6	7	8	9
55	4	1685	51	3	1530	4	34	1
44	2	1685	39	5	1490	5	27	0
45	2	1559	35	−1	1580	2	34	1
48	5	1705	45	5	1500	2	28	0
44	2	1723	44	5	1600	2	41	0
59	5	1700	47	−1	1570	3	39	1
64	5	1660	57	−1	1620	4	32	0
34	5	1681	33	−1	1410	2	22	0
37	3	1803	38	5	1560	2	23	1
54	2	1866	59	3	1590	2	49	0
49	3	1884	46	2	1710	2	25	0
63	3	1705	60	2	1580	3	27	0
48	2	1780	47	−1	1690	2	22	0
64	5	1801	55	3	1610	5	37	1
33	2	1795	45	3	1660	4	17	0
52	1	1669	47	3	1610	2	23	0
27	2	1708	24	2	1590	2	26	1
33	3	1691	32	3	1530	2	21	0
46	2	1825	47	2	1690	2	23	1
54	4	1760	57	5	1600	2	23	1
27	2	1949	−1	−1	1693	2	25	0
50	4	1685	−1	−1	1580	4	21	1
42	2	1806	−1	−1	1636	2	22	0
54	1	1905	46	3	1670	2	32	0
49	2	1739	42	5	1600	2	28	1
62	1	1736	63	−1	1570	3	22	0
34	3	1845	32	−1	1700	2	24	0
23	4	1868	24	−1	1740	2	19	1
36	1	1765	32	−1	1540	2	27	1
53	2	1736	−1	−1	1555	2	30	0
32	5	1741	−1	−1	1614	2	22	1
59	4	1720	56	5	1530	4	24	1
53	5	1871	50	−1	1690	3	25	0
55	5	1720	55	6	1590	4	21	1
62	3	1629	58	−1	1610	5	23	0
42	4	1624	38	−1	1670	2	22	1
50	6	1653	44	5	1690	3	35	1
37	2	1786	35	2	1550	2	21	1
51	2	1620	44	5	1650	2	30	0
25	3	1695	25	5	1540	2	19	0
54	3	1674	43	−1	1600	3	35	0
34	2	1864	31	3	1620	2	23	0
43	5	1643	35	5	1630	4	29	1
43	5	1705	41	3	1610	2	22	1
58	4	1736	50	3	1540	2	32	1
28	3	1691	23	−1	1610	2	23	0
45	4	1753	43	3	1630	2	21	1
47	2	1680	49	3	1530	4	20	1
57	3	1724	59	−1	1520	3	24	0
27	5	1710	−1	−1	1544	4	20	0
34	4	1638	38	−1	1570	4	33	1
57	2	1725	42	−1	1580	−1	52	1
27	5	1725	21	−1	1550	2	24	1
54	2	1630	−1	−1	1570	2	34	0
24	4	1810	−1	−1	1521	4	16	1
48	1	1774	42	3	1580	2	30	1
37	5	1771	35	5	1630	2	28	0
25	4	1815	26	−1	1650	2	20	0
57	3	1575	57	3	1640	3	20	0
40	2	1729	34	5	1650	3	26	0
61	3	1749	63	6	1520	1	21	1
25	4	1705	23	2	1620	2	24	0
32	4	1875	−1	−1	1744	2	22	1
37	1	1784	−1	−1	1647	2	22	1
45	4	1584	−1	−1	1615	4	29	1
24	2	1774	23	4	1680	2	22	1
47	5	1658	46	6	1670	4	24	0
44	4	1790	40	3	1620	2	24	0
52	3	1798	53	5	1570	2	25	0
45	2	1824	40	2	1660	2	23	1
20	4	1796	22	−1	1550	4	19	0
60	2	1725	60	1	1590	2	21	1
36	1	1685	32	5	1620	2	25	0
25	4	1769	24	5	1560	5	18	1
25	3	1749	28	−1	1670	2	21	1
35	6	1716	40	2	1650	2	17	0
35	5	1664	−1	−1	1539	4	22	1
49	5	1773	48	−1	1470	3	21	1
33	5	1760	33	5	1580	2	20	0
50	2	1725	49	3	1670	3	23	0
63	−1	1645	64	−1	1520	3	28	0
57	2	1694	55	2	1620	3	24	1
41	3	1851	41	3	1710	2	23	1
38	1	1691	38	2	1530	2	20	0
30	4	1880	31	3	1630	4	22	1
52	2	1835	52	−1	1720	4	30	1
51	4	1730	43	6	1570	4	22	1
46	4	1644	51	−1	1560	4	27	0
50	3	1723	47	3	1650	4	25	0
32	2	1758	−1	−1	1635	2	24	1
52	5	1718	32	3	1590	2	25	1
30	4	1723	33	−1	1590	4	22	1
33	2	1708	−1	−1	1566	3	21	0
20	4	1786	18	5	1590	4	19	1
32	4	1764	−1	−1	1662	−1	−1	1
51	5	1675	45	5	1550	4	25	1
64	5	1641	64	3	1570	4	30	1
44	5	1743	43	3	1560	4	25	1
40	2	1823	39	3	1630	2	23	0
59	4	1720	56	5	1530	4	24	1

Appendix B Tables

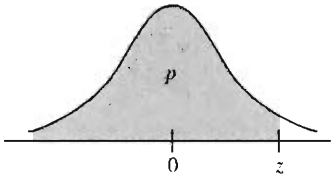


Table 1 Normal distribution function

The table gives the probability p that a normally distributed random variable Z with zero mean and unit variance is less than or equal to z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

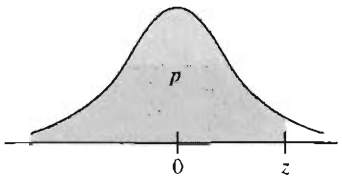


Table 2 Percentage points of the normal distribution

The table gives the values of z satisfying $P(Z \leq z) = p$ where Z is a normally distributed random variable with zero mean and unit variance.

p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.50	0.000	0.025	0.050	0.075	0.100	0.126	0.151	0.176	0.202	0.228
0.60	0.253	0.279	0.305	0.332	0.358	0.385	0.412	0.440	0.468	0.496
0.70	0.524	0.553	0.583	0.613	0.643	0.674	0.706	0.739	0.772	0.806
0.80	0.842	0.878	0.915	0.954	0.994	1.036	1.080	1.126	1.175	1.227
0.90	1.282	1.341	1.405	1.476	1.555					

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
0.99	2.326	2.366	2.409	2.457	2.512	2.576	2.652	2.748	2.878	3.090

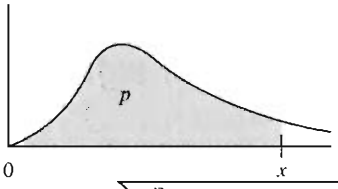


Table 3 Percentage points of the χ^2 distribution

The table gives the value of x satisfying $P(X \le x) = p$ where X is a χ^2 random variable with ν degrees of freedom.

ν \ p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.00004	0.0002	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892	62.883
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
45	24.311	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957	73.166
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
55	31.735	33.570	36.398	38.958	42.060	68.796	73.311	77.380	82.292	85.749
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
65	39.383	41.444	44.603	47.450	50.883	79.973	84.821	89.177	94.422	98.105
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
75	47.206	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393	110.286
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
85	55.170	57.634	61.389	64.749	68.777	102.079	107.522	112.393	118.236	122.325
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
95	63.250	65.898	69.925	73.520	77.818	113.038	118.752	123.858	129.973	134.247
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table 4 Critical values of the Wilcoxon signed ranks statistic

The table gives the lower tail critical values T_c of the statistic

$$T = \sum_{i=1}^{n} U_i R_i$$

where R_i denotes the rank of the magnitude of the ith observation in a sample of size n and $U_i = 1$ or 0 according as to whether this observation is positive or negative. The upper tail critical values are given by $\frac{1}{2}n(n+1) - T_c$. Since T is a discrete variable, exact significance levels cannot in general be achieved. The critical values given are those whose significance levels are nearest to those stated.

One tail	10%	5%	2.5%	1%	0.5%
Two tail	20%	10%	5%	2%	1%
n					
3	0				
4	1	0			
5	2	1	0		
6	4	2	1	0	
7	6	4	2	0	0
8	8	6	4	2	0
9	11	8	6	3	2
10	14	11	8	5	3
11	18	14	11	7	5
12	22	17	14	10	7
13	26	21	17	13	10
14	31	26	21	16	13
15	37	30	25	20	16
16	42	36	30	24	19
17	49	41	35	28	23
18	55	47	40	33	28
19	62	54	46	38	32
20	70	60	52	43	37
21	78	68	59	49	43
22	86	75	66	56	49
23	95	83	73	62	55
24	104	92	81	69	61
25	114	101	90	77	68
26	124	110	98	85	76
27	135	120	107	93	84
28	146	130	117	102	92
29	157	141	127	111	100
30	169	152	137	120	109
32	195	175	159	141	128
34	222	201	183	162	149
36	250	228	208	186	171
38	281	256	235	211	195
40	314	287	264	238	221
45	403	371	344	313	292
50	504	466	434	398	373

Table 5 Critical values of the Mann–Whitney statistic

The table gives the upper tail critical value u_c of the statistic

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}$$

where $Z_{ij} = 1$ if $X_i < Y_j$ and $Z_{ij} = 0$ if $X_i > Y_j$ given the independent samples $X_1, X_2, \ldots X_m$ and $Y_1, Y_2, \ldots Y_n$. The lower tail critical values are given by $mn - u_c$.

Since U is discrete, exact significance levels cannot in general be acheived. The critical values given are those whose significance levels are nearest to those stated.

One tail 0.5% Two tail 1%

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1																														
2													26	28	30	32	34	36	38	40	42	44	45	47	49	51	53	55	57	59
3							21	24	27	30	32	35	38	40	43	46	48	51	54	57	59	62	65	67	70	73	75	78	81	83
4					20	24	27	31	34	38	41	45	48	51	55	58	62	65	69	72	75	79	82	86	89	93	96	99	103	106
5				20	25	29	33	37	41	46	50	54	58	62	66	70	74	79	83	87	91	95	99	103	107	112	116	120	124	128
6				24	29	34	39	44	48	53	58	63	68	73	77	82	87	92	97	101	106	111	116	120	125	130	135	140	144	149
7			21	27	33	39	44	50	55	61	66	72	77	83	88	94	99	105	110	116	121	127	132	137	143	148	154	159	165	170
8			24	31	37	44	50	56	62	69	75	81	87	93	99	105	111	118	124	130	136	142	148	154	160	166	172	178	185	191
9			27	34	41	48	55	62	69	76	83	90	97	103	110	117	124	130	137	144	151	157	164	171	177	184	191	198	204	211
10			30	38	46	53	61	69	76	84	91	99	106	113	121	128	136	143	150	158	165	172	180	187	195	202	209	217	224	231
11			32	41	50	58	66	75	83	91	99	107	115	123	131	140	148	156	164	172	180	188	196	204	212	220	227	235	243	251
12			35	45	54	63	72	81	90	99	107	116	125	133	142	151	159	168	177	185	194	203	211	220	228	237	246	254	263	271
13		26	38	48	58	68	77	87	97	106	115	125	134	143	153	162	171	181	190	199	208	218	227	236	245	254	264	273	282	291
14		28	40	51	62	73	83	93	103	113	123	133	143	153	163	173	183	193	203	213	223	233	242	252	262	272	282	292	301	311
15		30	43	55	66	77	88	99	110	121	131	142	153	163	174	184	195	206	216	226	237	247	258	268	279	289	300	310	320	331
16		32	46	58	70	82	94	105	117	128	140	151	162	173	184	196	207	218	229	240	251	262	273	284	295	307	318	329	340	351
17		34	48	62	74	87	99	111	124	136	148	159	171	183	195	207	219	230	242	254	265	277	289	300	312	324	335	347	359	370
18		36	51	65	79	92	105	118	130	143	156	168	181	193	206	218	230	243	255	267	280	292	304	316	329	341	353	365	378	390
19		38	54	69	83	97	110	124	137	150	164	177	190	203	216	229	242	255	268	281	294	307	320	332	345	358	371	384	397	409
20		40	57	72	87	101	116	130	144	158	172	185	199	213	226	240	254	267	281	294	308	321	335	348	362	375	389	402	416	429
21		42	59	75	91	106	121	136	151	165	180	194	208	223	237	251	265	280	294	308	322	336	350	364	378	392	406	420	435	449
22		44	62	79	95	111	127	142	157	172	188	203	218	233	247	262	277	292	307	321	336	351	366	380	395	410	424	439	453	468
23		45	65	82	99	116	132	148	164	180	196	211	227	242	258	273	289	304	320	335	350	366	381	396	411	427	442	457	472	487
24		47	67	86	103	120	137	154	171	187	204	220	236	252	268	284	300	316	332	348	364	380	396	412	428	444	459	475	491	507
25		49	70	89	107	125	143	160	177	195	212	228	245	262	279	295	312	329	345	362	378	395	411	428	444	461	477	494	510	526
26		51	73	93	112	130	148	166	184	202	220	237	254	272	289	307	324	341	358	375	392	410	427	444	461	478	495	512	529	546
27		53	75	96	116	135	154	172	191	209	227	246	264	282	300	318	335	353	371	389	406	424	442	459	477	495	512	530	548	565
28		55	78	99	120	140	159	178	198	217	235	254	273	292	310	329	347	365	384	402	420	439	457	475	494	512	530	548	566	584
29		57	81	103	124	144	165	185	204	224	243	263	282	301	320	340	359	378	397	416	435	453	472	491	510	529	548	566	585	604
30		59	83	106	128	149	170	191	211	231	251	271	291	311	331	351	370	390	409	429	449	468	487	507	526	546	565	584	604	623

One tail 1% Two tail 2%

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1																														
2									18	20	22	24	26	28	30	31	33	35	37	39	41	43	44	46	48	50	52	54	56	56
3					15	18	21	23	26	29	31	34	36	39	42	44	47	50	52	55	57	60	63	65	68	70	73	76	78	81
4				16	20	23	26	30	33	36	40	43	46	50	53	56	60	63	66	70	73	76	79	83	86	89	93	96	99	103
5			15	20	24	28	32	36	40	44	48	52	56	60	64	68	72	76	80	84	88	92	96	100	104	108	112	116	120	124
6			18	23	28	33	37	42	47	51	56	61	65	70	75	79	84	89	93	98	102	107	112	116	121	126	130	135	139	144
7			21	26	32	37	43	48	53	59	64	69	75	80	85	91	96	101	106	112	117	122	128	133	138	143	149	154	159	164
8			23	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120	125	131	137	143	149	155	161	167	173	178	184
9		18	26	33	40	47	53	60	67	73	80	87	93	100	106	113	119	126	133	139	146	152	159	165	172	178	185	191	198	204
10		20	29	36	44	51	59	66	73	81	88	95	102	110	117	124	131	138	145	153	160	167	174	181	188	195	203	210	217	224
11		22	31	40	48	56	64	72	80	88	96	104	112	119	127	135	143	151	158	166	174	182	189	197	205	213	220	228	236	244
12		24	34	43	52	61	69	78	87	95	104	112	121	129	138	146	154	163	171	180	188	196	205	213	221	230	238	246	255	263
13		26	36	46	56	65	75	84	93	102	112	121	130	139	148	157	166	175	184	193	202	211	220	229	238	247	256	265	274	283
14		28	39	50	60	70	80	90	100	110	119	129	139	148	158	168	177	187	197	206	216	225	235	245	254	264	273	283	292	302
15		30	42	53	64	75	85	96	106	117	127	138	148	158	168	179	189	199	209	220	230	240	250	260	270	281	291	301	311	321
16		31	44	56	68	79	91	102	113	124	135	146	157	168	179	190	200	211	222	233	244	254	265	276	287	297	308	319	330	340
17		33	47	60	72	84	96	108	119	131	143	154	166	177	189	200	212	223	235	246	257	269	280	292	303	314	326	337	348	360
18		35	50	63	76	89	101	114	126	138	151	163	175	187	199	211	223	235	247	259	271	283	295	307	319	331	343	355	367	379
19		37	52	66	80	93	106	120	133	145	158	171	184	197	209	222	235	247	260	273	285	298	310	323	335	348	360	373	385	398
20		39	55	70	84	98	112	125	139	153	166	180	193	206	220	233	246	259	273	286	299	312	325	338	351	365	378	391	404	417
21		41	57	73	88	102	117	131	146	160	174	188	202	216	230	244	257	271	285	299	313	326	340	354	368	381	395	409	422	436
22		43	60	76	92	107	122	137	152	167	182	196	211	225	240	254	269	283	298	312	326	341	355	369	384	398	412	427	441	455
23		44	63	79	96	112	128	143	159	174	189	205	220	235	250	265	280	295	310	325	340	355	370	385	400	415	430	444	459	474
24		46	65	83	100	116	133	149	165	181	197	213	229	245	260	276	292	307	323	338	354	369	385	400	416	431	447	462	478	493
25		48	68	86	104	121	138	155	172	188	205	221	238	254	270	287	303	319	335	351	368	384	400	416	432	448	464	480	496	512
26		50	70	89	108	126	143	161	178	195	213	230	247	264	281	297	314	331	348	365	381	398	415	431	448	465	481	498	515	531
27		52	73	93	112	130	149	167	185	203	220	238	256	273	291	308	326	343	360	378	395	412	430	447	464	481	499	516	533	550
28		54	76	96	116	135	154	173	191	210	228	246	265	283	301	319	337	355	373	391	409	427	444	462	480	498	516	534	551	569
29		56	78	99	120	139	159	178	198	217	236	255	274	292	311	330	348	367	385	404	422	441	459	478	496	515	533	551	570	588
30		58	81	103	124	144	164	184	204	224	244	263	283	302	321	340	360	379	398	417	436	455	474	493	512	531	550	569	588	607

One tail 2.5% Two tail 5%

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1																				20	21	22	23	24	25	26	27	28	29	30
2					10	12	14	16	18	19	21	23	25	26	28	30	32	34	35	37	39	41	42	44	46	48	50	51	53	55
3				12	15	17	19	22	24	27	29	32	34	37	39	42	44	47	49	52	54	57	59	62	64	67	69	72	74	76
4			12	15	18	22	25	28	31	34	37	40	44	47	50	53	56	59	62	66	69	72	75	78	81	84	87	91	94	97
5		10	15	18	22	26	30	34	38	41	45	49	53	56	60	64	68	72	75	79	83	87	90	94	98	102	105	109	113	117
6		12	17	22	26	31	35	40	44	48	53	57	62	66	70	75	79	84	88	92	97	101	106	110	114	119	123	128	132	136
7		14	19	25	30	35	40	45	50	55	60	66	71	76	81	86	91	96	101	106	111	116	121	126	131	136	141	146	151	156
8		16	22	28	34	40	45	51	57	62	68	74	79	85	91	96	102	108	113	119	124	130	136	141	147	152	158	164	169	175
9		18	24	31	38	44	50	57	63	69	76	82	88	94	101	107	113	119	126	132	138	144	151	157	163	169	175	182	188	194
10		19	27	34	41	48	55	62	69	76	83	90	97	104	111	117	124	131	138	145	152	158	165	172	179	186	193	199	206	213
11		21	29	37	45	53	60	68	76	83	91	98	106	113	121	128	135	143	150	158	165	173	180	187	195	202	210	217	224	232
12		23	32	40	49	57	66	74	82	90	98	106	114	122	130	139	147	155	163	171	179	187	195	203	211	219	227	235	243	251
13		25	34	44	53	62	71	79	88	97	106	114	123	132	140	149	158	166	175	183	192	201	209	218	226	235	244	252	261	269
14		26	37	47	56	66	76	85	94	104	113	122	132	141	150	159	169	178	187	196	205	215	224	233	242	251	261	270	279	288
15		28	39	50	60	70	81	91	101	111	121	130	140	150	160	170	180	189	199	209	219	229	238	248	258	268	277	287	297	307
16		30	42	53	64	75	86	96	107	117	128	139	149	159	170	180	191	201	211	222	232	243	253	263	274	284	294	305	315	325
17		32	44	56	68	79	91	102	113	124	135	147	158	169	180	191	202	213	224	235	246	256	267	278	289	300	311	322	333	344
18		34	47	59	72	84	96	108	119	131	143	155	166	178	189	201	213	224	236	247	259	270	282	293	305	316	328	339	351	362
19		35	49	62	75	88	101	113	126	138	150	163	175	187	199	211	224	236	248	260	272	284	296	308	320	333	345	357	369	381
20	20	37	52	66	79	92	106	119	132	145	158	171	183	196	209	222	235	247	260	273	285	298	311	323	336	349	361	374	387	399
21	21	39	54	69	83	97	111	124	138	152	165	179	192	205	219	232	246	259	272	285	299	312	325	338	352	365	378	391	404	418
22	22	41	57	72	87	101	116	130	144	158	173	187	201	215	229	243	256	270	284	298	312	326	340	353	367	381	395	409	422	436
23	23	42	59	75	90	106	122	136	151	165	180	195	209	224	238	253	267	282	296	311	325	340	354	368	383	397	411	426	440	454
24	24	44	62	78	94	110	126	141	157	172	187	203	218	233	248	263	278	293	308	323	338	353	368	383	398	413	428	443	458	473
25	25	46	64	81	98	114	131	147	163	179	195	211	226	242	258	274	289	305	320	336	352	367	383	398	414	429	445	460	476	491
26	26	48	67	84	102	119	136	152	169	186	202	219	235	251	268	284	300	316	333	349	365	381	397	413	429	445	461	477	493	510
27	27	50	69	87	105	123	141	158	175	193	210	227	244	261	277	294	311	328	345	361	378	395	411	428	445	461	478	495	511	528
28	28	51	72	91	109	128	146	164	182	199	217	235	252	270	287	305	322	339	357	374	391	409	426	443	460	477	495	512	529	546
29	29	53	74	94	113	132	151	169	188	206	224	243	261	279	297	315	333	351	369	387	404	422	440	458	476	493	511	529	547	564
30	30	55	76	97	117	136	156	175	194	213	232	251	269	288	307	325	344	362	381	399	418	436	454	473	491	510	528	546	564	583

One tail 5% Two tail 10%

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1										10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	28	29
2			6	8	10	12	13	15	17	18	20	22	23	25	27	28	30	32	33	35	37	39	40	42	44	45	47	49	50	52
3		6	9	11	14	16	18	21	23	25	28	30	32	35	37	39	42	44	46	49	51	53	56	58	60	63	65	67	70	72
4		8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53	56	59	62	65	68	71	74	77	80	83	85	88	91
5		10	14	17	21	25	28	32	35	39	43	46	50	53	57	61	64	68	71	75	78	82	86	89	93	96	100	103	107	111
6		12	16	20	25	29	33	37	42	46	50	54	58	63	67	71	75	79	83	88	92	96	100	104	109	113	117	121	125	129
7		13	18	23	28	33	38	43	48	52	57	62	67	72	76	81	86	91	96	100	105	110	115	119	124	129	134	138	143	148
8		15	21	26	32	37	43	48	54	59	64	70	75	81	66	91	97	102	108	113	118	124	129	134	140	145	150	156	161	167
9		17	23	29	35	42	48	54	60	66	72	78	84	90	96	102	108	114	120	125	131	137	143	149	155	161	167	173	179	185
10	10	18	25	32	39	46	52	59	66	72	79	86	92	99	105	112	118	125	131	138	145	151	158	164	171	177	184	190	197	203
11	11	20	28	35	43	50	57	64	72	79	86	93	100	108	115	122	129	136	143	150	158	165	172	179	186	193	200	207	214	221
12	12	22	30	38	46	54	62	70	78	86	93	101	109	117	124	132	140	147	155	163	171	178	186	194	201	209	217	224	232	240
13	13	23	32	41	50	58	67	75	84	92	100	109	117	125	134	142	150	159	167	175	183	192	200	208	217	225	233	241	250	258
14	14	25	35	44	53	63	72	81	90	99	108	117	125	134	143	152	161	170	179	188	196	205	214	223	232	241	249	258	267	276
15	15	27	37	47	57	67	76	86	96	105	115	124	134	143	153	162	172	181	190	200	209	219	228	238	247	256	266	275	285	294
16	16	28	39	50	61	71	81	91	102	112	122	132	142	152	162	172	182	192	202	212	222	232	242	252	262	272	282	292	302	312
17	17	30	42	53	64	75	86	97	108	118	129	140	150	161	172	182	193	203	214	225	235	246	256	267	277	288	298	309	319	330
18	18	32	44	56	68	79	91	102	114	125	136	147	159	170	181	192	203	215	226	237	248	259	270	281	292	303	315	326	337	348
19	19	33	46	59	71	83	96	108	120	131	143	155	167	179	190	202	214	226	237	249	261	272	284	296	307	319	331	342	354	366
20	20	35	49	62	75	88	100	113	125	138	150	163	175	188	200	212	225	237	249	261	274	286	298	310	232	335	347	359	371	384
21	21	37	51	65	78	92	105	118	131	145	158	171	183	196	209	222	235	248	261	274	286	299	312	325	338	350	363	376	389	402
22	22	39	53	68	82	96	110	124	137	151	165	178	192	205	219	232	246	259	272	286	299	313	326	339	353	366	379	393	406	419
23	23	40	56	71	86	100	115	129	143	158	172	186	200	214	228	242	256	270	284	298	312	326	340	354	368	382	396	409	423	437
24	24	42	58	74	89	104	119	134	149	164	179	194	208	223	238	252	267	281	296	310	325	339	354	368	383	397	412	426	441	455
25	25	44	60	77	93	109	124	140	155	171	186	201	217	232	247	262	277	292	307	323	338	353	368	383	398	413	428	443	458	473
26	26	45	63	80	96	113	129	145	161	177	193	209	225	241	256	272	288	303	319	335	350	366	382	397	413	428	444	460	475	491
27	27	47	65	83	100	117	134	150	167	184	200	217	233	249	266	282	298	315	331	347	363	379	396	412	428	444	460	476	492	509
28	28	49	67	85	103	121	138	156	173	190	207	224	241	258	275	292	309	326	342	359	376	393	409	426	443	460	476	493	510	526
29	28	50	70	88	107	125	143	161	179	197	214	232	250	267	285	302	319	337	354	371	389	406	423	441	458	475	492	510	527	544
30	29	52	72	91	111	129	148	167	185	203	221	240	258	276	294	312	330	348	366	384	402	419	437	455	473	491	509	526	544	562

A

abstract reasoning, gender differences (Act. 3.11) 50
algorithms
DiceFreq 63, 125
FindU 43, 120
ProbU 43, 120
Weldon 107

B

'before and after' tests 18–19, 34
binomial distribution 57, 58
(Act. 4.2) 58
(Acts. 4.8–4.10) 66–9
birthdays (Act. 5.1) 84
checklist 69
and Poisson distribution (Acts. 5.3, 5.4) 86–7
birthdays
(Act. 5.1) 84
(Act. 5.4) 86–7
body language (Act. 4.18) 80–81
body weight, night and morning, (Acts. 3.3, 3.4) 39, 40
'British Young People in the 80s' 5
response rates 25
sample 22, 23

C

chi-squared distribution 60–61
(Act. 4.4) 61
(Act. 4.5) 62
(Act. 4.6) 63–4
(Acts. 4.7–4.10) 64–9
(Act. 4.11) 71
table 143
test of association 82
(Acts. 4.16–4.18) 76–81
(Act. 5.4) 87
cigarettes, tar content (Act. 3.11) 51
closed questions 27
cluster sampling **23**
coding 9–10
computers 39, 47
confounding factors **14**, 15–16
(Act. 2.5) 14–15
(Acts. 2.6, 2.7) 16–18
constraints **124**
contingency tables **76**, 82
continuous data, measurement 11
continuous probability distribution 99
control group **15**, 16
controlled variables **13**
correlation *see* chi-squared distribution, test of association

D

data collection 17–32
(Act. 1.2) 6
data measurement 9–11
(Act. 2.2) 12
data sources 6
degrees of freedom **60–61**, 82
dental treatments, comparison (Act. 2.4) 14
dependent variables (response variables) **12**, 13
deviation (Act. 4.3) 59–60
dice experiment *see* Weldon's experiment
discrete data, measurement 11
discrete probability distribution 978
discrete simulation (Act. 6.2) 98
distribution free tests **52**
see also non-parametric tests
double blind experiments **16**
drawing pins experiments 56
(Acts. 4.9, 4.10) 67–9
drink–driving association (Act. 2.4) 14

E

euthanasia survey (Act. 2.18) 30
examination anxiety, treatment (Act. 3.11) 50
expected frequencies **58**–9, 66–7, 78, 82
experimenter effect 16
experiments **9**
(Act. 2.1) 9
confounding factors **14**, 15–16
(Act. 2.5) 14–15
(Acts. 2.6, 2.7) 16–18
designing (Act. 2.8) **19**
types 18–19
variables **12**–13
(Acts. 2.3, 2.4) 13–14
explanatory variables **12**, 13
extraneous variables **13**, 16

F

fire retardants experiment (Act. 2.4) 14
friction experiment, modelling data 74–6

G

goodness of fit, measuring **59**, 61–2
(Act. 4.3) 59–60
(Act. 4.16) 78
Government Statistics 6
graph paper *see* normal probability graph paper; Poisson graph paper

H

hair/eye colour relationship (Act. 4.17) 80
handedness, and spatial ability (Act. 3.11) 51
hay fever experiments 5–6
health survey
(Act. 2.16) 26–7
(Act. 2.17) 29
HEIGHT dataset 12, 23–4, 138–41
humans, experiments on 9, 18, 19
hypnosis, effectiveness (Act. 3.11) 50
hypothesis testing (Acts. 5.7, 5.8) 89–90

I

independent samples **18**, 41
(Act. 3.5) 41–2
(Act. 3.10) 49–50
independent variables (explanatory variables) **12**, 13
instruments 38
interval scales **10–11**
IQ scores experiment (Act. 2.4) 14

J

job satisfaction survey (Act. 4.17) 79–80

L

leading questions 27–8
lung cancer 1, 9

M

Mann-Whitney *U* test **44**, 47–8, 52
tables 47, 145–7
see also separation between data sets (*U*)
marginal totals **78**
market research surveys 5, 22
matched groups 18, 34
maturation 16
mean deviation (Act. 4.15) 76
measles vaccine experiments 56–7
(Act. 4.17) 79
measurement of data 9–11
(Act. 2.2) 12
memory experiments 34–6
(Act. 3.1) 36
(Act. 3.4) 40
modelling 1
Monte Carlo simulation 95

N

nominal scale of measurement 10, 11
non-parametric tests 44, **48**, **52**
non-random sampling 20–23
(Act. 2.10) 21
(Act. 2.11) 22
(Act. 2.12) 23
non-respondents 24
(Act. 2.14) 25
normal distribution 57, 69
(Act. 4.11) 69–71
checklist 76
tables 141–2
see also normal probability graph paper
normal probability graph paper 72, 74–6
(Act. 4.15) 76
(Acts. 4.12–4.14) 72–4
normal simulation (Act. 6.3) 99

O

observed frequencies **58**, **78**–9
occupations, classification 10
old people, effects of exercise
(Act. 2.5) 15
(Act. 3.11) 51
open questions 27
order effects 16
ordinal scale of measurement 10, 11

P

parametric tests **52**
Pearson, Karl 57, 61–2
pedestrian crossing, usage simulation 95
pesticide usage (survey) 5
pilot surveys 26, 27
placebo effect **15**
(Act. 2.7) 17–18
Poisson distribution 83–94
checklist 94
model for hypothesis testing
(Acts. 5.7–5.10) 89–91
model for random events 85, 87–9
(Acts. 5.3–5.5) 86–7
(Act. 5.6) 89
notation 88
Poisson graph paper 92–4
(Act. 5.11) 92
population **8**
(Act. 2.1) 8
post office arrivals (Act. 5.10, 5.11) 91–4
practice effects **16**
project duration, simulation 95
pulse rates, modelling data
(Act. 4.13) 72–3

Q

questionnaires 26, 27–9
(Act. 2.16) 26–7
queueing simulations
(Act. 6.1) 97
(Act. 6.3) 99
quota samples 22–3
(Act. 2.12) 23

R

random events
(Act. 5.2) 84–5
Poisson distribution as
model 85, 87–9
(Act. 5.5) 87
random numbers, using 20, 98
random sampling **19**–20
randomisation 16, 18, 20, 40
ratio scales **10**–11
reaction times
(Act.3.4) 40
modelling data
(Act. 4.11) 69–71
(Act. 4.12) 72
repeated measures design **18**
see also 'before and after' tests
representative samples **19**
response rates **24**–5
response variables **12**, 13
robustness of methods **52**
running shoes, comparison
(Act. 2.5) 15, 16
running times, altitude effects
(Act. 3.3) 40

S

sample statistics **8**
samples **8**
(Act. 2.1) 8
(Act. 2.13) 23
cluster **23**
independent **18**, 41
(Act. 3.5) 41–2
(Act. 3.10) 49–50
quota 22–3
(Act. 2.12) 23
representative **19**
simple random **20**
(Act. 2.9) 20
stratified 21–2
(Act. 2.11) 22
systematic **20**–21
(Act. 2.10) 21
young people's survey 5
separation between data
sets (*U*) 42, 44, 46
(Act. 3.7) 43–4
(Acts. 3.9–3.11) 48–5
formula (Act. 3.8) 46–7
tables 47, 145–7
sexual discrimination survey
(Act. 2.19) 30–31
shoppers' survey 5
shrimp eggs, hatching rates
(Act. 3.7) 45–6, 47
sign test (Act. 3.1) 36
significance of results 38–9
(Act. 3.6) 43–4
simple random samples **20**
(Act. 2.9) 20
simple random sampling **19**–20
simulations 56–7, 95–9
chi-squared test (Act. 4.6) 63–4
Weldon's experiment (Act. 4.1) 57–8
Smarties, colours (Act. 4.8) 66–7
smoking 1, 9
(Act. 1.1) 1–2
by vegetarians (Act. 4.16) 76–9
and infant birthweight (Act. 3.9) 49
social class, indicators 10
soil acidity survey 42–3, 47
(Act. 3.5) 41–2
(Act. 3.8) 46–7
(Act. 3.9) 48–9
split polls **30**
(Acts. 2.18, 2.19) 30–31
standard deviation (Act. 4.15) 76
statistical investigations 5–6
personal 3–4
(Act. 2.20) 31
stem-and-leaf diagrams (Act. 3.1) 36
stereoscopic vision (Act. 3.4) 40–41
stratified samples 21–2
(Act. 2.11) 22
surveys **9**
(Act. 2.15) 25–6
(Act. 2.1) 9
pilot 26, 27
swimming times
(Act. 3.3) 39
(Act. 3.9) 48
systematic samples **20**–21
(Act. 2.10) 21

T

t test 52
taste test (Act. 2.4) 14
teaching methods, comparisons
(Act. 2.7) 15, 17
TOWNS dataset 12
twin studies 18

U

U see separation between data sets

V

variables **12**–13
(Acts. 2.3, 2.4) 13–14
vegetarians, and smoking
(Act. 4.16) 76–9
vitamin tablets, effectiveness
(Act. 2.5) 15, 16
(Act. 2.7) 17
volunteer samples 15

W

watches, where worn (Act. 4.17) 79
Weldon's experiment 57, 58, 61–2
modelling data
(Act. 4.2) 58–9
(Act. 4.3) 59–60
(Act. 4.4) 61
(Act. 4.5) 62
simulation (Act. 4.1) 57–8
Wilcoxon signed ranks
test **36**, 38–9, 41, 52
(Act. 3.2) 37–8
(Acts. 3.3, 3.4) 39–41
(Act. 3.11) 50–51
table 144
word processing (Act. 3.11) 50–51

Y

young people *see* 'British Young People in the 80s'

STATISTICS

Nuffield Advanced Mathematics focuses on the differing needs of individuals by offering students the opportunity to follow their own interests, to apply mathematics in solving real problems and to study a range of new and exciting options.

As part of their extra work, A level Mathematics students choose one option from this range which includes: *Mechanics 1, Mechanics 2, Statistics, Complex numbers and numerical methods, Discrete mathematics, Surfaces, The history of mathematics* and *Mathematics, music and art.* Students can take Further Mathematics at AS or A level by taking a further three or six options.

The Nuffield Advanced Mathematics series also includes:
Book 1 0582 09311 2
Book 2 0582 09314 7
Book 3 0582 09975 7
Book 4 0582 09976 5
Book 5 0582 09977 3
Teacher's notes
Resources file
Mathematics reader

ISBN 0-582-09985-4

9 780582 099852

LONGMAN