# Choosing and Using Statistics

## A Biologist's Guide

**Calvin Dytham**

*Department of Biology, University of York*

**Second Edition**

Blackwell
Publishing

# CHOOSING AND USING STATISTICS
A BIOLOGIST'S GUIDE

# Choosing and Using Statistics

## A Biologist's Guide

**Calvin Dytham**

*Department of Biology, University of York*

**Second Edition**

Blackwell
Publishing

350 Main Street, Malden, MA 02148-5018, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton South, Melbourne, Victoria 3053, Australia
Kurfürstendamm 57, 10707 Berlin, Germany

First published 1999
Second edition 2003

*Library of Congress Cataloging-in-Publication Data*

Dytham, Calvin.
Choosing and using statistics: a biologist's guide/
Calvin Dytham.—2nd ed.
p. cm.
Includes index.
ISBN 1-40510-243-8 (pbk.: alk. paper)
1. Biometry. I. Title.

QH323.5.D98 2003
570′.1′5195—dc21
2002074760

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

Set in $9\frac{1}{2}$/12pt Berling
by Graphicraft Limited, Hong Kong
Printed and bound in the United Kingdom
by TJ International Ltd, Padstow, Cornwall

For further information on
Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com

# Contents

# Preface

My aim was to produce a statistics book with two characteristics: to assume that the reader is using a computer to analyse data and to contain absolutely no equations.

This book is a handbook for biologists who want to process their data through a statistical package on the computer, to select the most appropriate methods and extract the important information from the, often confusing, output that is produced. It is aimed, primarily, at undergraduates and masters students in the biological sciences who have to use statistics in practical classes and projects. Such users of statistics do not have to understand exactly how the test works or how to do the actual calculations and these things are not covered in this book as there are more than enough books that do that already. What is important is that the right statistical test is used and the right inferences made from the output of the test. An extensive key to statistical tests is included for the former and the bulk of the book is made up of descriptions of how to carry out the tests to address the latter.

In several years of teaching statistics to biology students it is clear to me that most students do not really care how or why the test works. They do care a great deal that they are using an appropriate test and interpreting the results properly. I think that this is a fair aim to have for occasional users of statistics. Of course, anyone going on to use statistics frequently should become familiar with the way that calculations manipulate the data to produce the output as this will give a better understanding of the test.

If this book has a message it is this: *think about the statistics* ***before*** *you collect the data!* So many times I have seen rather distraught students unable to analyse their precious data because the experimental design they used was inappropriate. On such occasions I try to find a compromise test that will make the best of a bad job but this often leads to a weaker conclusion than might have been possible if more forethought had been applied from the outset. There is no doubt that if experiments or sampling strategies are designed with the statistics in mind better science will result.

Statistics are often seen by students as the 'thing you must do to data at the end'. Please try to avoid falling into this trap yourself. 'Thought' experiments producing dummy data are a good way to try out experimental designs and are much less labour intensive than real ones!

Although there are no equations in this book I am afraid there was no way to totally avoid statistical jargon. To ease the pain somewhat, an extensive glossary and key to symbols is included. So when you are navigating your way through the key to choosing a test you should look up any words you do not understand.

In this book I have given extensive instructions for the use of three widely available software packages: SPSS, MINITAB and Excel. However, the key to choosing a statistical test is not at all package specific so if you use a software package other than the three I focus on, or if you a using a calculator, you will still be able to get a good deal out of this book.

If every sample gave the same result there would be no need for statistics. All aspects of biology seem to be filled with variation. It is statistics that can be used to penetrate the haze of experimental error and the inherent variability of the natural world to reach the underlying causes and processes at work. So, try not to hate statistics, they are merely a tool that, when used wisely and properly, can make the life of a biologist much simpler and give conclusions a sound basis.

## The second edition

In the 4 years since I wrote the first edition of this book there have been several new versions of the software produced. This was inevitable, and has been the main driver for the production of a second edition using more recent versions of the software. There are few changes to the structure and message of the book from the first edition. I have removed the mini-reviews of other statistics books and feel that anyone moving on from this book should consider moving to either Zar's *Biostatistical Analysis* or Sokal and Rohlf's *Biometry*. Both are large, explain how to calculate statistical tests and are more comprehensive than this volume.

I have received many comments about the first edition and I am grateful for the many suggestions on how to improve the text and coverage. Requests to add further statistical packages have been the most common suggestion for change. But, as there was surprisingly little consensus on the packages to add, I decided to restrict the coverage, as before, to SPSS, MINITAB and Excel. I apologize to all those users of Systat, Genstat, SAS, Statistica, S/S-plus/R and GLIM who were hoping their package would be included.

I have expanded some sections from the first edition. This includes a simple table to encapsulate the key, coverage of the *G*-test, more ANOVA designs, logistic regression and some other regression techniques. There are also more examples of multivariate techniques. I have tried to do this without shifting the audience and suggest that anyone needing to use a range of multivariate techniques should look elsewhere for guidance.

## How to use this book

This is definitely not a book that should be read from cover to cover. It is a book to refer to when you need assistance with statistical analysis, either when choosing an appropriate test or when carrying it out. The basics of statistical analysis and experimental design are covered briefly, but these sections are intended mostly as a revision aid, or outline of some of the more important concepts. Other statistics books may help you choose those that are most appropriate for you if you want or need more details.

The heart of the book is the key in Chapter 3. The rest of the book hinges on the key, explaining how to carry out the tests, giving assistance with the statistical terms in the glossary or giving tips on the use of computers and packages.

### Packages used

MINITAB® version 13.10 and 13.32, *MINITAB inc.*
SPSS® version 10.1, *SPSS inc.*
Excel™ versions 97 and 2000, *Microsoft Corporation*.
Running on: Windows® versions 95 (release 2), 98, ME and 2000, *Microsoft Corporation*.

### Example data

In the spirit of dummy data collection, all example data used throughout this book have been fabricated. Any similarity to data alive or dead is purely coincidental.

## Acknowledgements for the first edition

Thanks to Sheena McNamee for support during the writing process, to Andrea Gillmeister and two anonymous reviewers for commenting on an early version of the manuscript and to Terry Crawford, Jo Dunn, David Murrell and Josephine Pithon for recommending and lending various books. Thanks also to Ian Sherman and Susan Sternberg at Blackwells and to many of my colleagues who told me that the general idea of a book like this was a sound one. Finally, I would especially like to thank the students at the University of York, UK, who brought me the problems that provided the inspiration for this book.

## Acknowledgements for the second edition

Thanks to all the many people who contacted me with suggestions and comments about the first edition. I hope you can see that many of the corrections and improvements have come directly from you. Five anonymous reviewers

provided many useful comments about the proposal for a second edition. Thanks to Sarah Shannon, Cee Brandson, Katrina McCallum and many others at Blackwells for seeing this book through and especially for producing a second superb and striking cover. S'Albufera Natural Parc and Nick Riddiford provided a very convenient bolt-hole for writing. Once again, I give special thanks to Sheena and to my colleagues, PhD students and undergraduate students at the University of York. Finally, thanks to everyone on the MRes EEM course over the last 4 years.

Calvin Dytham
York, 1998 and 2002

# Eight steps to successful data analysis

# 1

This is a very simple sequence that, it you follow it, will integrate the statistics you use into the process of scientific investigation. As I make clear here, statistical tests should be considered *very early* in the process and not left until the end.

**1** Decide what you are interested in.

**2** Formulate a hypothesis or several hypotheses (see Chapters 2 and 3 for guidance).

**3** Design the experiment, manipulation or sampling routine that will allow you to test the hypotheses (see Chapters 2 and 4 for some hints on how to go about this).

**4** *Collect dummy data* (i.e. make up approximate values based on what you expect to obtain). The collection of 'dummy data' may seem strange but it will convert the proposed experimental design or sampling routine into something more tangible. The process can often expose flaws or weaknesses in the data collection routine that will save a huge amount of time and effort.

**5** Use the key (presented in Chapter 3) to guide you towards the appropriate test or tests.

**6** Carry out the test(s) using the dummy data. (Chapters 6–9 will show you how to input the data, use the statistical packages and interpret the output.)

**7** If there are problems go back to step 3 (or 2), otherwise proceed to the collection of the real data.

**8** Carry out the test(s) using the real data. Report the findings and/or return to step 2.

I implore you to use this sequence. I have seen countless students who have spent a long time and a lot of effort collecting data only to find that the experimental or sampling design was not quite right. The test they are forced to use is much less powerful than one they could have used with only a slight change in the experimental design. This sort of experience tends to turn people away from statistics and become 'scared' of them. This is a great shame as statistics are a hugely useful and vital tool in science.

The rest of the book follows this eight step process but you should use it for guidance and advice when you become unsure of what to do.

# 2 The basics

The aim of this section is to introduce, in rather broad terms, some of the recurring concepts of data collection and analysis. Everything introduced here is covered at greater length in later chapters and certainly in the many statistics text books that aim to introduce statistical theory and experimental design to scientists.

The key to statistical tests in the next chapter assumes that you are familiar with most of the basic concepts introduced here.

## Observations

These are the raw material of statistics and can include anything recorded as part of an investigation. They can be on any scale from a simple 'raining or not raining' dichotomy to a very sophisticated and precise analysis of nutrient concentrations. The type of observations recorded will have a great bearing on the type of statistical tests that are appropriate.

Observations can be simply divided into three types: *categorical* where the observations can be in a limited number of categories that have no obvious scale (e.g. 'oak', 'ash', 'elm'); *discrete* where there is a real scale but not all values are possible (e.g. 'number of eggs in a nest' or 'number of species in a sample'); and *continuous* where any value is theoretically possible but is restricted by the measuring device (e.g. lengths, concentrations).

Different types of observations are considered in more detail in Chapter 5.

## Hypothesis testing

The cornerstone of scientific analysis is hypothesis testing. The concept is rather simple: almost every time a statistical test is carried out it is testing the probability that a hypothesis is correct. If the probability is small then the hypothesis is deemed to be untrue and it is rejected in favour of an alternative. This is done in what seems to be a rather upside down way as the test is always of what is called the null hypothesis rather than the more interesting hypothesis. The null hypothesis is the hypothesis that nothing is going on (it is often labelled as $H_0$). For example, if the weights of bulbs for two cultivars of daffodils were being investigated, the null hypothesis would be that there is nothing going on: 'the weights of the two groups of bulbs are the same'. A statistical test is carried out to find out how likely that null hypothesis is to be true. If we decide to reject the null hypothesis we must accept the alternative, more interesting hypothesis ($H_1$) that: 'the weights of bulbs for the two cultivars are different'.

## *P*-values

The *P*-value is the bottom line of most statistical tests. It is simply the probability that the hypothesis being tested is true. So if a *P*-value is given as 0.06, that indicates that the hypothesis has a 6% chance of being true. In biology it is usual to take a value of 0.05 or 5% as the critical level for the rejection of a hypothesis. This means that providing a hypothesis has a less than one in 20 chance of being true, we reject it. As it is the null hypothesis that is nearly always being tested we are always looking for low *P*-values to reject this hypothesis and accept the more interesting alternative hypothesis.

Clearly the smaller the *P*-value the more confident we can be in the conclusions drawn from it. A *P*-value of 0.0001 indicates that the chance of the hypothesis being tested being true is one in 10 000 and this is much more convincing than a marginal $P = 0.049$.

*P*-values and the types of errors that are implicitly accepted by their use are considered further in Chapter 4.

## Sampling

Observations have to be collected in some way. This process of data acquisition is called sampling. Although there are almost as many different methods that can be used for sampling as there are possible things to sample, there are some general rules. One of the most obvious is that more observations are usually better than few. Balanced sampling is also important (i.e. when comparing two groups take the same number of observations from each group).

Most statistical tests assume that samples are taken at random. This sounds easy but is actually quite difficult to achieve. For example, if you are sampling beetles from pitfall traps the sample may seem totally random but in fact is quite biased towards those species that move around the most. Another common bias is to chose a point at random and then measure the nearest individual to that point assuming that this will produce a random sample. It will not be random at all as individuals at the edges of clumps are more likely to be selected than those in the middle. There are methods available to reduce problems associated with non-random sampling but the first step is to be aware of the problem.

A further assumption of sampling is that individuals are either only measured once or they are all sampled on several occasions. This assumption is often violated if, for example, the same site is visited on two occasions and the same individuals or clones are inadvertently remeasured.

The sets of observations collected are called variables. A variable can be almost anything it is possible to record as long as different individuals can be assigned different values.

Some of the problems of sampling are considered in Chapter 4.

## Experiments

In biology many investigations use experiments of some sort. An experiment occurs if anything is altered or controlled by the investigator. For example, an investigation into the effect of fertilizer on plant growth will use a control plot (or several control plots) where there is no fertilizer added and then one or more plots where fertilizer has been added at known concentrations set by the investigators. In this way the effect of fertilizer can be determined by comparison of the different concentrations of fertilizer. The condition being controlled (e.g. fertilizer) is usually called a factor and the different levels used called treatments or factor levels (e.g. concentrations of fertilizer). The design of this experiment will be determined by the hypothesis or hypotheses being investigated. If the effect of the fertilizer on a particular plant is of interest then, perhaps, a range of different soil types might be used with and without fertilizer. If the effect on plants in general is of interest then an experiment using a variety of plants is required, either in isolation or together. If the optimum fertilizer treatment is required then a range of concentrations will be applied and a cost–benefit analysis carried out.

More details and strategies for experimental design are considered in Chapter 4.

## Statistics

In general, statistics are the results of manipulation of observations to produce a single or small number of results. There are various categories of statistics depending on the type of summary required. Here I divide statistics into four categories.

### Descriptive statistics

The simplest statistics are summaries of data sets. Simple summary statistics are easy to understand but should not be overlooked. These are not usually considered to be statistics but are in fact extremely useful for data investigation. The most widely used are measures of the ‘location’ of a set of numbers such as the mean or median. Then there are measures of the ‘spread’ of the data, such as the standard deviation. The choice of appropriate descriptive statistics and the best way of displaying the results are considered in Chapters 5 and 6.

## Tests of difference

A familiar question in any field of investigation is going to be something like '*is this group different from that group?*'. A question of this kind can then be turned into a null hypothesis with a form: '*this group and that group are not different*'. To answer this question, and test the null hypothesis, a statistical test of difference is required. There are many tests that all seem to answer the same type of question but each is appropriate when certain types of data are being considered. After the simple comparison of two groups there are extensions to comparisons of more than two groups and then to tests involving more than one way of dividing the individuals into groups. For example, individuals could be assigned to two groups by sex and also into groups depending on whether they had been given a drug or not. This could be considered as four groups or as what is known as a factorial test where there are two factors 'sex' and 'drug' with all combinations of the levels of the two factors being measured in some way. Factorial designs can become very complicated but they are very powerful and can expose subtleties in the way the factors interact that can never be found through investigation of the data using one factor at a time.

Tests of difference can also be used to compare variables with known distributions. These can be statistical distributions or derived from theory. Chapter 7 considers tests of difference in detail.

## Tests of relationships

Another familiar question that arises in scientific investigation is in the form '*is A associated with B?*'. For example 'is fat intake related to blood pressure?'. This type of question should then be turned into a null hypothesis that '*A is not associated with B*' and then tested using one of a variety of statistical tests. As with tests of difference there are many tests that seem to address the same type of problem, but again each is appropriate for different types of data.

Tests of relationships fall into two groups called correlation and regression depending on the type of hypothesis being investigated. Correlation is a test to measure the degree to which one set of data varies with another — it *does not* imply that there is any 'cause' and 'effect' relationship. Regression is used to fit a relationship between two variables such that one can be predicted from the other. This *does* imply a 'cause' and 'effect' relationship or at least an implication that one of the variables is a 'response' in some way. So in the investigation of fat intake and blood pressure a strong positive correlation between the two shows a relationship but does not show cause and effect. If there were a significant positive regression line, this would imply that blood pressure can be predicted using fat intake *or*, if the regression uses the fat intake as the 'response', that fat intake can be predicted from blood pressure.

There are many additional techniques that can be employed to consider the relationships between more than two sets of data. Tests of relationships are described in Chapter 8.

## Tests for data investigation

A whole range of tests is available to help investigators explore large data sets. Unlike the tests considered above, data investigation need not have a hypothesis for testing. For example, in a study of the morphology of fish there may be many fin measurements from a range of species and sites that offer far too many potential hypotheses for investigation. In this case the application of a multivariate technique may show up relationships between individuals, help assign unknown specimens to categories or just suggest which hypotheses are worth further consideration.

A few of the many different techniques available are considered in Chapter 9.

# Choosing a test: a key

# 3

I hope that you are not reading this section with your data already collected and the experiment or sampling programme ‘finished’. If you have finished collecting your data I strongly advise you to approach your next experiment or survey in a different way. As you will see below, I hope that you will be using this key *before* you start collecting real data.

## Remember — eight steps to successful data analysis

1 Decide what you are interested in.
2 Formulate a hypothesis or hypotheses.
3 Design the experiment, manipulation or sampling routine.
4 Collect dummy data. Make up approximate values based on what you expect.
5 *Use the key here to decide on the appropriate test or tests*.
6 Carry out the test(s) using the dummy data.
7 If there are problems go back to step 3 (or 2); otherwise collect the real data.
8 Carry out the test(s) using the real data.

## The art of choosing a test

It may be a surprising revelation, but choosing a test is not an exact science. There is nearly always scope for considerable choice and many decisions will be made based on personal judgements, experience with similar problems, or just plain hunches. There are many circumstances where there are several ways that the data could be analysed and yet each of the possible tests could be justified.

A common tendency is to force the data from your experiment into a test you are familiar with even if it is not the best method. Look around for different tests that may be more appropriate to the hypothesis you are testing. In this way you will expand your statistical repertoire and add power to your future experiments.

## A key to assist in your choice of statistical test

Starting at point 1 move through the key following the path that best describes your data. If you are unsure about any of the terms used then consult the glossary or the relevant sections of the next two chapters. This is not a true dichotomous key and at several points there are more than two routes or end points.

There may be several end points appropriate to your data that result from this key. For example you may wish to know the correct display method for the data and then the correct measure of dispersion to use. If this is the case, go through the key twice. All the tests and techniques mentioned in the key are described in later chapters.

Italics indicate instructions about what you should do.

Numbers in brackets indicate that the point in the key is something of a compromise destination.

There are several points where rather arbitrary numbers are used to determine which path you should take. For example, I use 30 different observations as the arbitrary level at which to split continuous and discontinuous data. If your data set falls close to this level you should not feel constrained to take one path if you feel more comfortable with the other.

---

1 Testing a clear hypothesis and associated null hypothesis (e.g. $H_1$ = blood glucose level is related to age and $H_0$ = blood glucose is not related to age). 25

Not testing any hypothesis but simply want to present, summarize or explore data. 2

2 Methods to summarize and display the data required. 3

Data exploration for the purpose of understanding and getting a feel for the data or perhaps to help with formulation of hypotheses. For example, you may wish to find possible groups within the data (e.g. 10 morphological variables have been taken from a large number of carabid beetles; the multivariate test may establish whether they can be divided into separate taxa). 60

3 There is only one collected variable under consideration (e.g. the only variable measured is brain volume although it may have been measured from several different populations). 4

There is more than one variable (e.g. you have measured the number of algae per millilitre *and* the water pH in the same sample). 24

4 The data are discrete; there are fewer than 30 different values (e.g. number of species in a sample). 5

The data are continuous; there are more than 29 different values (e.g. bee wing length measured to the nearest 0.01 mm). 16

(Note: the distinction between the above is rather arbitrary.)

5 There is only one group or sample (e.g. all measurements taken from the same river on the same day). 6

There is more than one group or sample (e.g. you have measured the number of antenna segments in a species of beetle and have divided the sample according to sex to give two groups). 15

6 A graphical representation of the data is required. 7

A numerical summary or description of the data required. 11

7 A display of the whole distribution is required. 8

A crude display of the position and spread of data is required — *use a box and whisker display to show medians, range and interquartile range, p. 46 (also known as a box plot)*.

8 Values have real meaning (e.g. number of mammals caught per night). 10

Values are arbitrary labels that have no real sequence (e.g. different soil type classifications in an area of forest). 9

9 There are fewer than 10 different values or classifications — *draw a pie chart, p. 49. Ensure that each segment is labelled clearly and that adjacent shading patterns are as distinct as possible. Avoid using 3D or shadow effects, dark shading or colour. Do not add the proportion in figures to the 'piece' of the pie as this information is redundant.*

There are 10 or more different values or classifications — *amalgamate values until there are fewer than 10 or divide the sample to produce two sets each with fewer than 10 values. Ten is a level above which it is difficult to distinguish different sections of the pie or to have sufficiently distinct shading patterns.*

10 There are more than 20 different values — *amalgamate values to produce around 12 classes (almost certainly done automatically by your package) and draw a histogram, p. 47. Give the classes on the* x-*axis, frequency of occurrence (number of times the value occurs) on the* y-*axis, and no gaps between bars. Do not use 3D or shadow effects.*

There are 20 or fewer different values — *draw a bar chart, p. 47. Each value should be represented on the* x-*axis. If there are few classes, extend the range to include values not in the data set at either side; give frequency of occurrence (number of times the value occurs) on the* y-*axis. Gaps should appear between bars, unless the variable is clearly supposed to be continuous; do not use 3D or shadow effects.*

11 Measure of position (mean is the one used most commonly). 12

Measure of dispersion or spread (standard deviation and confidence intervals are the most commonly used). 13

Measure of symmetry or shape of the distribution. 14

(Note: you will probably want to go for at least one measure of position and another of spread in most cases.)

12 Variable is definitely discrete and usually restricted to integer values smaller than 30 (e.g. number of eggs in a clutch) — *calculate the median, p. 50.*
Variable should be continuous but has only a few different values due to accuracy of measurement (e.g. bone length measured to the nearest centimetre) — *calculate the mean, p. 49.*
If you are particularly interested in the most commonly occurring response — *calculate the mode, p. 50, in addition to either the mean or median.*

13 Very rough measure of spread is required — *calculate the range, p. 51 (note that this measure is very biased by sample size and is rarely a useful statistic).*
You are particularly interested in the highest and/or lowest values — *calculate the range, p. 51.*
Variable should be continuous but has only a few values due to accuracy of measurement — *calculate the standard deviation, p. 52.*
Variable is discrete or has an unusual distribution — *calculate the interquartile range, p. 51.*

14 Variable should be continuous but has only a few values due to accuracy of measurement — *calculate the skew* ($g_1$)*, p. 53.*
Observations are discrete or if you have already calculated the interquartile range and the median — *the relative size of the interquartile range above and below the median provides a measure of the symmetry of the data.*

15 You have not established the appropriate technique for a single sample — *go back to 6 to find the appropriate techniques for each group. You should find the same is correct for each sample or group.* (6)
The samples can be displayed separately — *go back to 7 and choose the appropriate style. So that direct comparisons can be made, be sure to use the same scales (both* x*-axis and* y*-axis) for each graph. Be warned that packages will often adjust scales for you. If this happens you must force the scales to be the same.* (7)
The samples are to be displayed together on the same graph — *use a chart with a box plot for each sample and the* x*-axis representing the sample number, p. 57. Make sure there is a clear space between each box plot.*

16 There is only a data set from one group or sample. 17
The data have been collected from more than one group or sample (e.g. you have measured the mass of each individual of a single species of vole from one sample and have divided the sample according to sex). 23

17 A graphical representation of the data is required. 18
A numerical summary or descriptive statistics are required. 19

18 Display of the whole distribution is required — *group to produce around 12–20 classes and draw a histogram, p. 47 (probably done automatically by your package). Use classes on the* x*-axis, frequency of occurrence (number of times the value occurs within the class) on the* y*-axis, and no gaps between bars, no 3D or shadow effects. Even-sized classes are much easier for a reader to interpret. Data with an unusual distribution (e.g. there are some extremely high values well away from most of the observations) may require transformation before the histogram is attempted.*
A crude display of position and spread of the data is required — *the 'error bar' type of display is unusual for a single sample but common for several samples. There is a symbol representing the mean and a vertical line representing the range of either the 95% confidence interval or the standard deviation, p. 57.*

19 Measure of position (mean is the most common). 20
Measure of dispersion (spread). 21
Measure of symmetry or shape of the distribution. 22
You wish to determine whether the data are normally distributed — *carry out a Kolmogorov–Smirnov test, p. 77, an Anderson–Darling test, p. 79 or a chi-square goodness of fit, p. 68.*
(Note: you probably require one of each of the above for a full summary of the data.)

20 Unless the variable is definitely discrete or is known to have an odd distribution (e.g. not symmetrical) — *calculate the mean, p. 49.*
If the data are known to be discrete or the data set is to be compared with other, discrete data with fewer possible values — *calculate the median, p. 50.*
If you are particularly interested in the most commonly occurring value — *calculate the mode, p. 50,* ***in addition to*** *the mean or median.*

21 If the data are continuous and approximately normally distributed and you require an estimate of the spread of data — *calculate the standard deviation (SD), p. 52.*
(Note: standard deviation is the square root of variance and is measured in the same units as the original data.)
If you have previously calculated the mean and require an estimate of the range of possible values for the mean — *calculate 95% confidence limits for the mean, p. 52 (a.k.a. 95% confidence interval or 95% CI).*
A very rough measure of spread is required — *calculate the range, p. 51. (Note that this measure is very biased by sample size and is rarely a useful statistic in large samples.)*
If you have a special interest in the highest and or lowest values in the sample — *calculate the range, p. 51.*