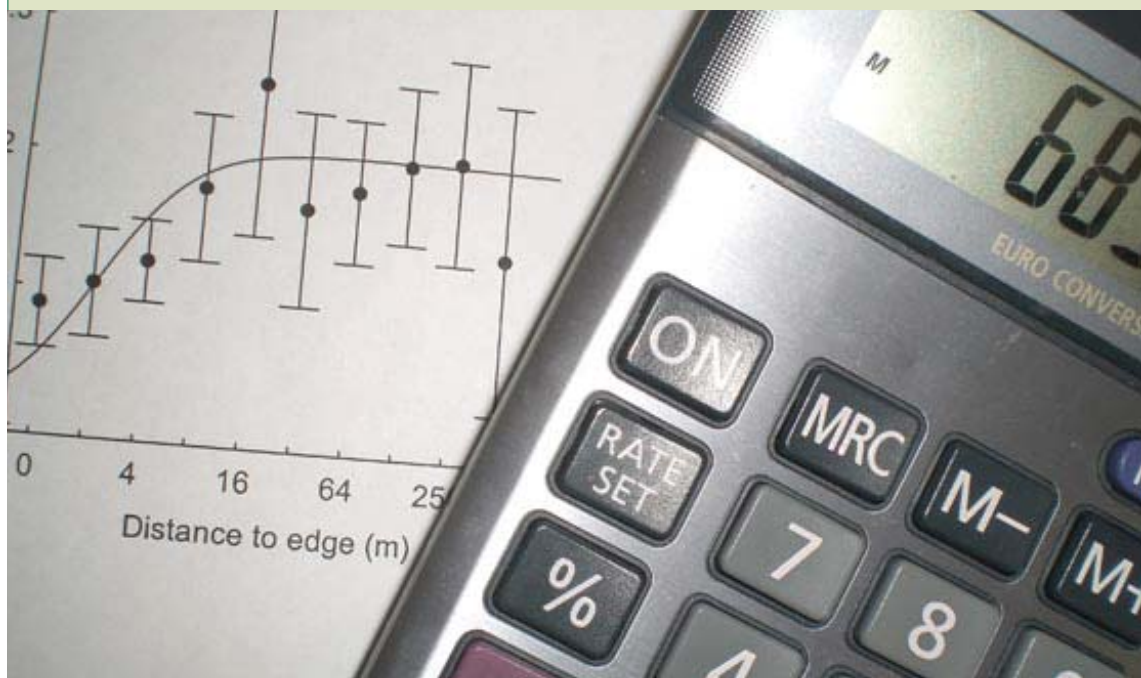




TROPICAL BIOLOGY ASSOCIATION

A simple guide to Minitab



A simple guide to minitab is designed to familiarise you with some of the statistical tests available to you on a software programme on the TBA laptops. It provides instructions to carry out tests using Minitab v13.32 for Windows and to help you to interpret the output. It does not tell you which test is right for your analysis, or what assumptions should be met before the test is valid. This information is provided elsewhere e.g. in some of the books to be found in the TBA travelling library and you will need to refer to this before analysing data.

How you use this manual is up to you. You may wish to sit down in front of a computer with Minitab running and work your way through the book, plugging examples into the computer as you go. Or you may prefer to use the guide as a reference manual, looking up specific tests as you need them. In either case, we hope you find this guide takes some of the fear out of statistics: computer analysis is a tool like any other, and will take a lot of the hard work out of statistics once you feel you are in charge!

For any queries concerning this document please contact:

Tropical Biology Association

Department of Zoology

Downing Street, Cambridge

CB2 3EJ

United Kingdom

Tel: +44 (0) 1223 336619

e-mail: tba@tropical-biology.org

© **Tropical Biology Association 2008**

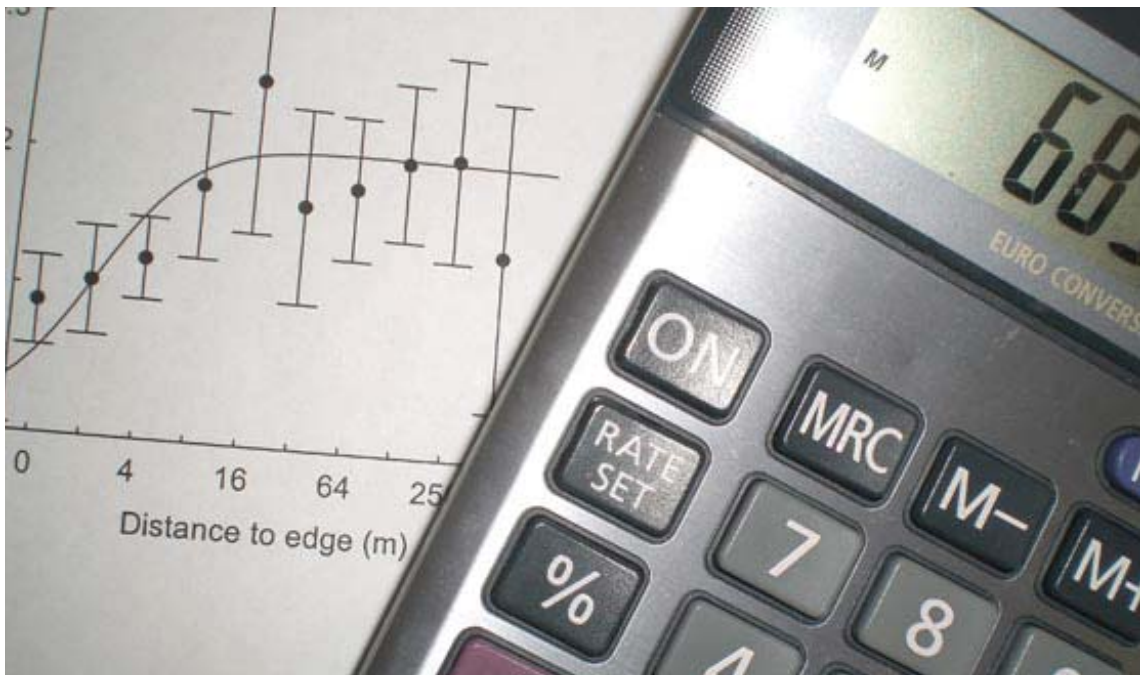
Acknowledgement

This booklet was adapted for use on the TBA courses with the kind permission of Dr Su Engstrand, University of Highlands & Islands.



TROPICAL BIOLOGY ASSOCIATION

A simple guide to Minitab



CONTENTS

GETTING STARTED	5
1.1 NOTATION IN THIS MANUAL	5
1.2 MINITAB'S WINDOWS	5
1.3 ENTERING DATA	5
1.4 SAVING YOUR WORK	6
1.5 MANIPULATING DATA	7
1.5.1 Copying and pasting data	7
1.5.2 Manipulating data	7
1.5.3 Stacking and un-stacking data	8
SUMMARISING DATA	10
2.1 GRAPHING DATA	10
2.1.1. Graph>Histogram	10
2.1.2. Looking for relationships: Graph>Plot	11
STATISTICAL TESTING	13
3.1 TESTING ASSUMPTIONS	13
3.1.1 Testing for normality	13
3.1.2 Homogeneity of variance	14
3.2 PARAMETRIC TESTS	14
3.2.1 Independent samples t-test	14
3.2. Matched pairs (or paired samples) t-test	16
3.2.3 Analysis of variance	17
3.2.3.1 One-way ANOVA	17
3.2.3.2 Two-way ANOVA	19
3.2.4 Pearson product-moment correlation	21
3.2.5 Linear regression	21
3.3 NON-PARAMETRIC TESTS	23
3.3.1. Mann-Whitney test	23
3.3.2 Wilcoxon signed rank test	23
3.3.3 Kruskal-Wallis test	24
3.3.4 Spearman rank correlation	25
3.3.5 Chi-square test	25
3.3.5.1. Contingency table	25
3.3.5.2 Goodness-of-fit test	25

GETTING STARTED

Minitab is a statistical package, which is widely used by biologists, biomedical and medical scientists. It is very powerful and, for the type of project you undertake on a TBA course, you are unlikely to need to do any analysis that Minitab cannot cope with. However, it can seem a bit 'user un-friendly' to start with. **DON'T PANIC!** Just bear this in mind, try not to be put off and you should find yourself equipped with a powerful tool to help you interpret your data.

You will find Minitab installed on all the TBA computers. This manual is written to correspond to version 13, don't worry that we are using an older version and there are slight differences in how your screen looks. After starting your PC's you can access Minitab from the **desktop icon**.

1.1 NOTATION IN THIS MANUAL

Almost all of the tests you will require Minitab to perform can be accessed from the menu which should appear across the top of the screen as soon as you open Minitab. This is much simpler than typing in commands (as we used to do in days gone by). Hopefully, you will all be familiar with using computer menus by now (from packages such as Word or Excel). In this manual, **bold type** is used to show the name of the menu command to be chosen. Choosing a command will often lead to a second choice from a subsequent menu and this second choice will be shown after the symbol '>'. For example, the instructions to display descriptive statistics would be written as follows: **Stat>Basic Statistics>Display Descriptive Statistics**. This would lead you to a screen as shown in Figure 1 on the next page.

1.2 MINITAB'S WINDOWS

When you open up a Minitab file, you are presented with two windows: the **Worksheet** window, where you will enter and store your data, and the **Session** window, which shows the results of your manipulations

and analysis (Figure 1). You can arrange your screen to show both windows simultaneously, or to show either window across the full screen. Do this by clicking on the square buttons in the top right hand corner of each window. Alternatively, choose the Session or Data window from the list given under **Windows** in the top menu bar. You will also find windows for any graphs produced appearing in this list.

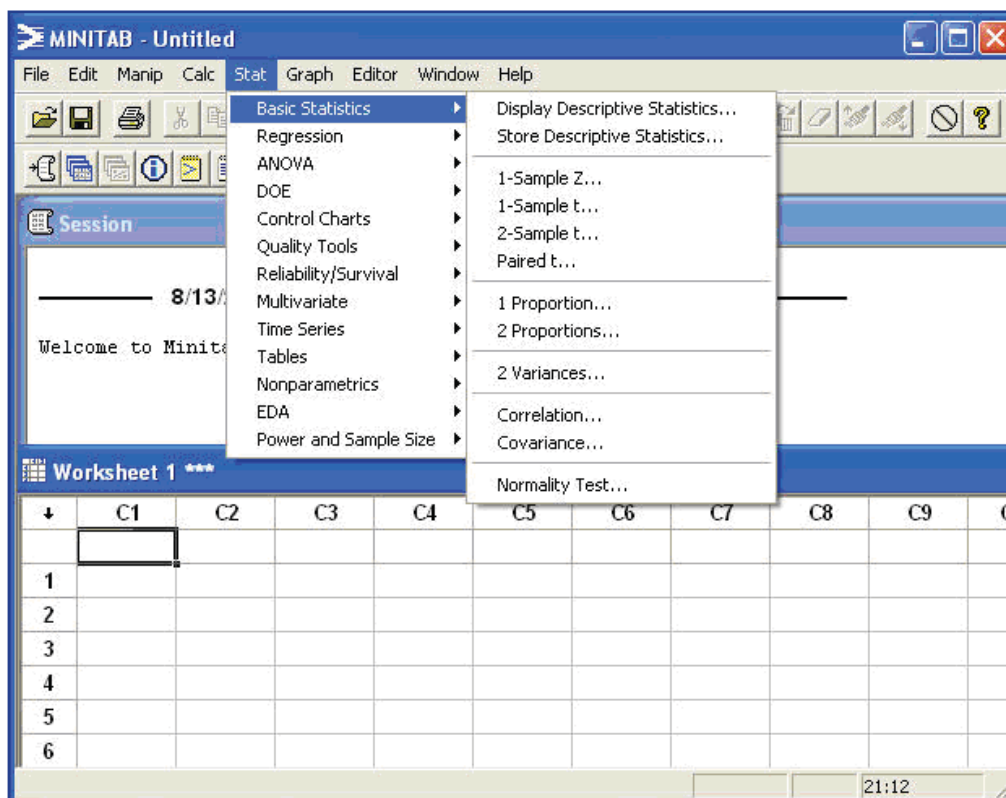
1.3 ENTERING DATA

Minitab presents you with a blank worksheet, with columns, labelled C1, C2...etc. across the top, and row numbers shown down the side. The most common way to enter data is to put the names of each **variable** you have measured at the top of each column. Use the mouse to position the cursor immediately below the labels C1, C2 etc., then click and enter the variable name. It is a good idea to make these as specific as possible, so that you can remember exactly what data you want to choose for your analysis. You can then enter data for a single variable, by moving vertically down a column, or you can enter the data for every subject you have measured, variable by variable,

Producing a data set

In most cases, you will want to enter data directly into Minitab, though there may be occasions when you wish to import data from another package. When entering data directly, you should first think hard about how you wish the data to be arranged. This manual will give instructions as to how the data should best be entered in Minitab for each test you are to perform, so **read the appropriate section before you start typing** in your data. Most participants choose to enter their data set in Excel and you can cut and paste directly from that programme into Minitab; however, beware of problems of text vs. numeric data (see next tip box).

Figure 1
Screen shot showing how to choose the commands for descriptive statistics.



moving horizontally across the screen. Use the mouse or the cursor keys ($\downarrow \leftarrow \rightarrow$) to move around the screen.

Figure 2 shows data for biometrics measured on ten swallows. The data for each of 7 **variables** ('site' to 'mass') are listed for each bird, so that you can read off the values for any one individual by reading left to right across the spreadsheet. Note that the first variable ('site') is labelled by Minitab as C1-T. The 'T' in this label indicates that the data in this column is text rather than numeric.

1.4 SAVING YOUR WORK

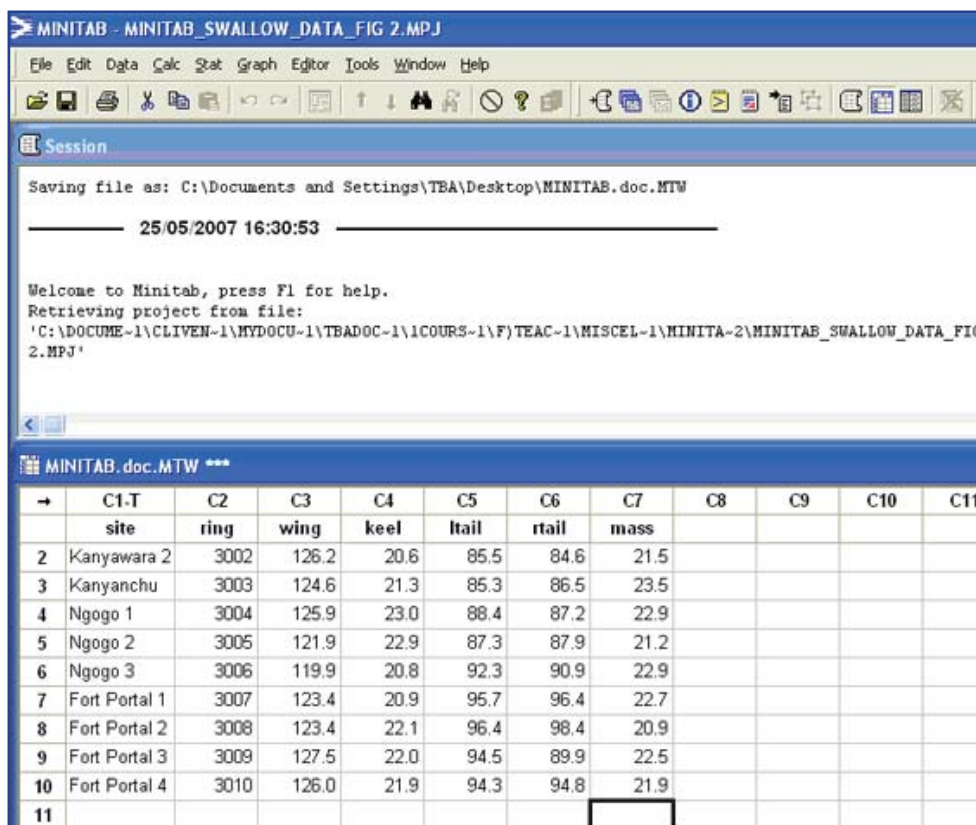
As with all work on the computer, it is a good idea to save your work at frequent intervals to avoid losses when the computer crashes ...or the power is cut off

...or you delete all your work by mistake... all three do happen quite regularly! To save your work for the first time, choose **'File> Save Project As'**. Saving a project saves all open worksheets, graphs and the session window together in a single file. You will be asked to give details of where you want to save the worksheet and what you want to call it. You should normally choose to save the file to a folder you create on the desktop of your laptop.

Name your file in the **'File name'** box – try to use a sensible naming procedure which might include the topic, date, or file contents. Click on **Save** on the right hand side of the window. You should get very used to this procedure. Just remember that the first time you save any document you have to name it.

EXAMPLE of trouble shooting - Text versus numeric data

Common difficulties with text data (i.e. letters instead of numbers). In Figure 2, the site names are text. But if, you wanted to enter numeric data, but you typed a letter into a cell by mistake, even if you then erase it, Minitab may act as though you have entered text. You are then likely to find that you cannot perform the stats that you want using this column. So if you find a 'T' in any column heading, when the data are really numeric, you must get rid of the problem by converting text to numeric data: first click anywhere in the session window to make it active. Choose **Editor>Enable Commands** and the flashing cursor should appear after the Minitab prompt **MTB>**. Type **'aton c1 c1'** if you wish to replace text data in c1 with numeric data in c1. (If your data are in c2, obviously change what you type accordingly). Press enter.



MINITAB - MINITAB_SWALLOW_DATA_FIG 2.MPJ

File Edit Data Calc Stat Graph Editor Tools Window Help

Session

Saving file as: C:\Documents and Settings\TBA\Desktop\MINITAB.doc.MTW

25/05/2007 16:30:53

Welcome to Minitab, press F1 for help.
Retrieving project from file:
'C:\DOCUME~1\CLIVEN~1\HYDOCU~1\TBADOC~1\1COURS~1\F\TEAC~1\MISCEL~1\MINITA~2\MINITAB_SWALLOW_DATA_FIG 2.MPJ'

MINITAB.doc.MTW ***

	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
	site	ring	wing	keel	ltail	rtail	mass				
2	Kanyawara 2	3002	126.2	20.6	85.5	84.6	21.5				
3	Kanyanchu	3003	124.6	21.3	85.3	86.5	23.5				
4	Ngogo 1	3004	125.9	23.0	88.4	87.2	22.9				
5	Ngogo 2	3005	121.9	22.9	87.3	87.9	21.2				
6	Ngogo 3	3006	119.9	20.8	92.3	90.9	22.9				
7	Fort Portal 1	3007	123.4	20.9	95.7	96.4	22.7				
8	Fort Portal 2	3008	123.4	22.1	96.4	98.4	20.9				
9	Fort Portal 3	3009	127.5	22.0	94.5	89.9	22.5				
10	Fort Portal 4	3010	126.0	21.9	94.3	94.8	21.9				
11											

Figure 2
Screenshot
showing
session window
and worksheet
containing
biometric
information for
10 swallows
captured in the
Kibale district.

When you are saving for the second or subsequent time, **File>Save Project** will automatically save the file under the name you gave originally, by replacing the old version, so you don't need to type in the file name again. Be careful: if you have closed any graphs since you last saved the project, the file will be replaced with the new project without these graphs. If you have created important graphs, it makes sense to save them separately in a file of their own. Click on the graph to make it active, then use the **File>Save Graph As** command.

When you wish to reopen your file, first open Minitab, then chose **File> Open Project**; choose your workspace and you should be able to click on the name of the file you want.

1.5 MANIPULATING DATA

1.5.1 Copying and pasting data

You can use the **Cut**, **Copy**, and **Paste** commands to bring data into Minitab from some other packages e.g. Excel, though you must be careful to get column headings etc. in the right place. You can also use these commands within Minitab to move data from one column to another, or as a short-cut in data entry. The commands are fairly self-explanatory.

1.5.2 Manipulating data

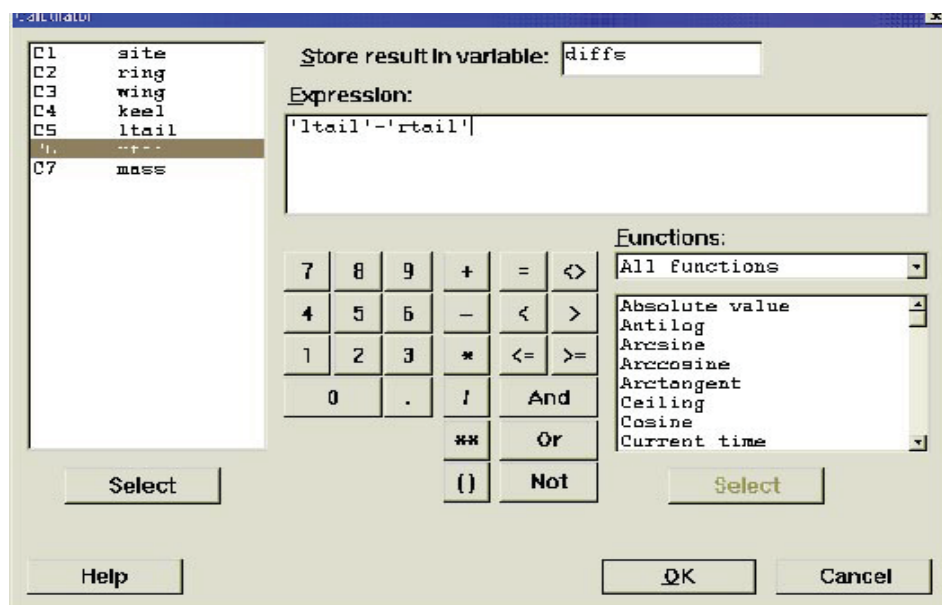
You may need to manipulate or transform data for a number of reasons. For this, use **Calc> Calculator**.

You are presented with a list of expressions (Figure 3), which allow you to perform several manipulations on the data. For example, the data in columns c5 and c6 above give the length of left and right tail feather for each swallow. You may be interested in the degree of asymmetry between these tail feathers, i.e. how much the left feather differs from the right. In this case, you could create a new variable, equal to the difference between two other variables. First you must decide where you want the new variable of differences to be stored and type it in the **Store result in Variable** box. Either choose an empty column for the new variable (e.g. c8) or else type a new name for the variable e.g. 'diffs' into the box, and Minitab will create a new variable of this name and put it in the first available empty column.

You would then need to type 'c5-c6' into the **expression** box. Alternatively, you can give Minitab the formula by clicking on the appropriate variables, as follows. First position the cursor where you wish the formula to end up (in this case, in the 'Expression' box) and click to make it active. Then move the mouse to the variables listed on the left and double click on the first variable you require. (Clicking once on the variable then choosing **Select** has the same effect but is slightly slower). You must then type in the function you require or select it from the options listed, in this example, type the minus sign, then double-click to select the second variable.

Figure 3

Using the Calculator command manipulates data = compute a new column ('diffs') of differences between the left and right tail feather length.



Other handy manipulations are those used to transform data when the raw data are not normally distributed. For example, if you wished to take logs to base 10 (log10), choose the appropriate function (in this case it is called 'LOGT') from those listed, select it or type it into the expressions box, then choose the variable to be transformed, to go inside the brackets. If we wanted to take log10 of left tail length, the dialog box would show **LOGT ('ltail')**.

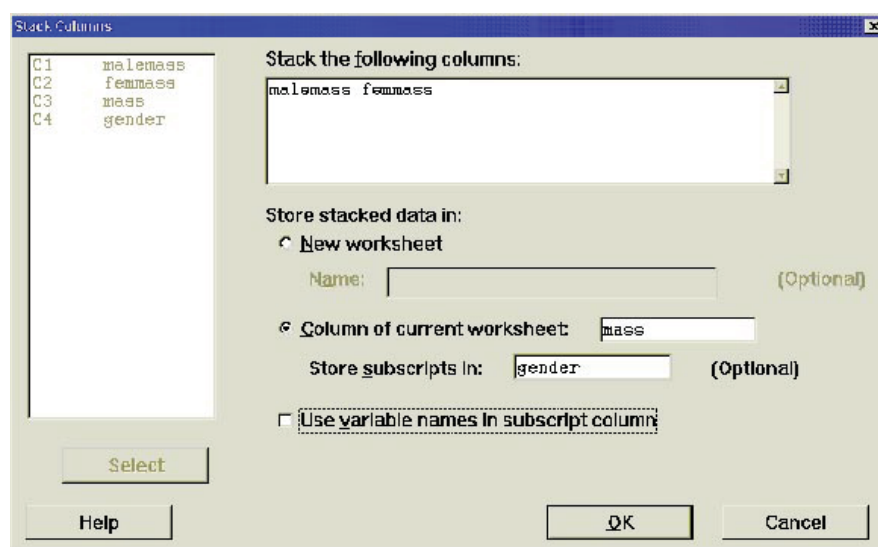
1.5.3 Stacking and un-stacking data

This is a very useful tool that will be helpful to you as you discover that certain tests require the data to be arranged in different formats. For example, let's say we have data on mass of both the male and female of 10 pairs of sunbirds. We might have entered the data as shown in columns 1 and 2 in Figure 5, with the mass of each male in c1 and that of the female

in c2. For certain functions, it is easier if the data on mass are presented in a single column. To do this, we use **Manip > Stack > Stack Columns**. Figure 4 shows how we choose the two variables and list them, then choose an empty column (c3) to store the data in. Alternatively, you can type an appropriate name into this column (in the example below we have used 'mass') and Minitab will enter the data into the first available empty column and label it with this name. We will also need to know which data came from the males and which from the females, so we choose to **store subscripts** by clicking on that box, and choose another column (c4, or label as 'gender') to store them in. By default, Minitab chooses to **use variable names in subscript column**. If you leave this box ticked, the new column of subscripts will contain text: 'malemass' for males and 'femmass' for females. As we have noted before, not all operations

Figure 4

Screenshot demonstrating the commands to stack two columns, containing data for male and female mass, into a single new column ('mass').



are possible with text (alphanumeric) data so you may find it more convenient to have the subscripts in numeric form. To do this, you should uncheck the box **to use variable names in subscript column**. Minitab will then print a '1' in c4 for all the male data (as malemass was listed first in the stacking command) and a '2' for all females (Figure 5).

A new column ('gender') will be created to show which data came from males and which from females. In this example, the command **to use variable names**

in subscript columns has been unchecked, so the new column gender will contain numerical values: 1 for all male data (as 'malemass' was entered first in the stacking command) and 2 for all female data.

Of course, you can do the reverse and unstack data by using **Manip> Unstack Columns** command (Figure 6). In this case, you would need to specify the subscripts by which the data will be unstacked (it would be c4 or 'gender' in our example). You can choose to **Name the columns containing the unstacked data**.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
	malemass	femmass	mass	gender											
1		21.6	24.5	21.6	1										
2		22.0	25.0	22.0	1										
3		15.2	23.6	15.2	1										
4		20.4	21.4	20.4	1										
5		19.4	23.3	19.4	1										
6		17.0	20.5	17.0	1										
7		18.0	24.5	18.0	1										
8		20.0	25.0	20.0	1										
9		19.6	23.6	19.6	1										
10		23.1	23.2	23.1	1										
11				24.5	2										
12				25.0	2										
13				23.6	2										
14				21.4	2										
15				23.3	2										
16				20.5	2										
17				24.5	2										
18				25.0	2										
19				23.6	2										
20				23.2	2										

Figure 5

Worksheet showing unstacked data for male and female sunbird mass (columns c1 and c2) and the same data stacked into a single column (c3) with subscripts indicating gender (c4), 1=male, 2 = female.

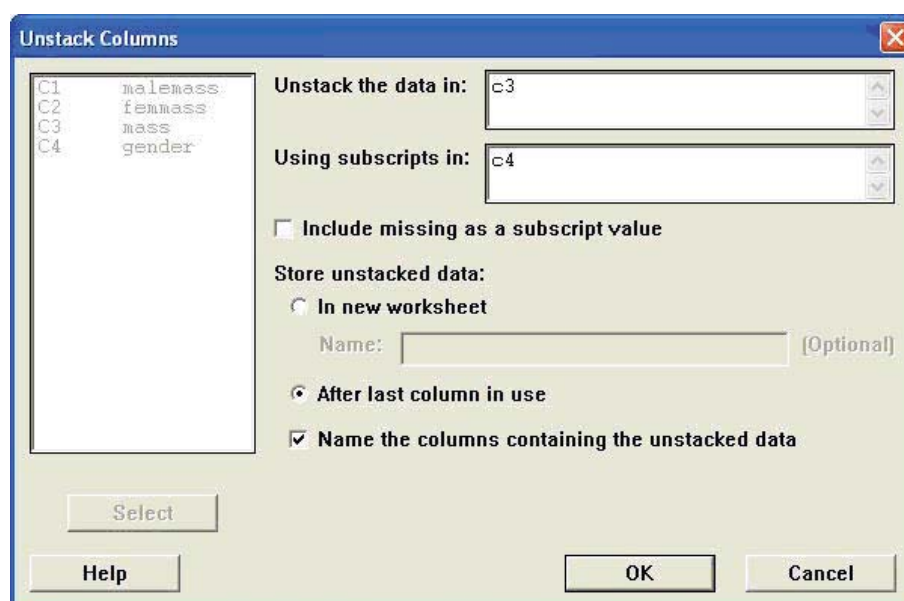


Figure 6

The commands to unstack data in a single column (c3) into new columns separated according to the subscripts in c4. The new columns will be labelled according to these subscripts (e.g. unstacking the mass data in c3 of Figure 5 would generate two new columns, mass_1 and mass_2).

SUMMARISING DATA

2.1 GRAPHING DATA

The best way to get a feel for the graphics which are available in Minitab is to click on **Graph** on the top menu bar, and to have a play around with some of the options. Many of the commands produce good quality graphs. Below are two types of graph which you might want to use.

2.1.1. Graph>Histogram

Click to select one or more of the variables to be

displayed e.g. malemass (Figure 7). This is very useful for looking at the shape of the distribution of a variable e.g. Figure 8.

You can alter a number of properties of the graph using the various buttons on the menu screen. For instance, Figure 9 displays the same data as Figure 8, but the number of bars has been changed to 10 by clicking the **Options** button, and changing the **number of intervals**.

Figure 7

The comand screen to produce two histograms, one for 'malemass' and a second for 'femmass'.

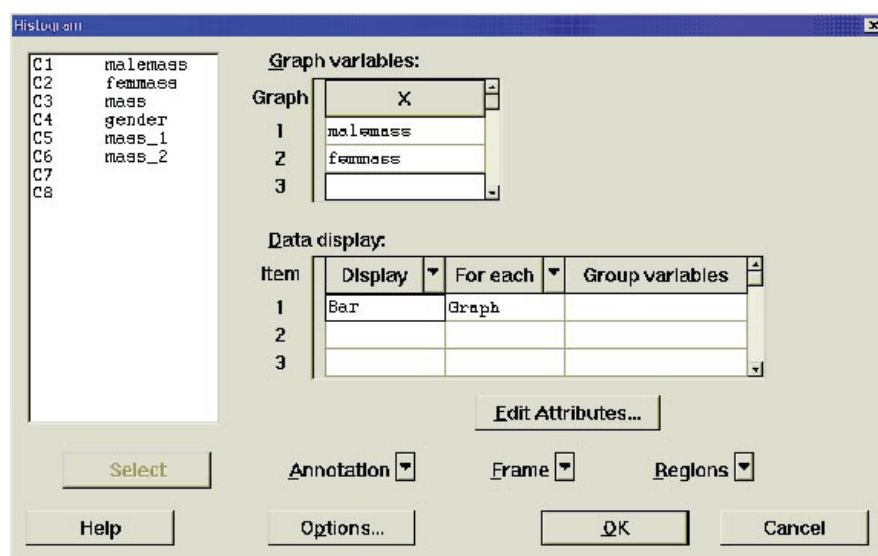
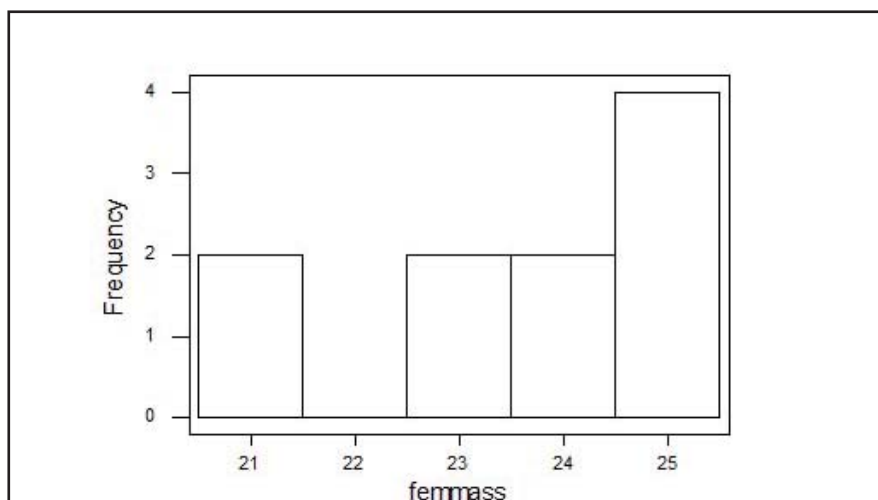


Figure 8

Frequency histogram for mass for the female sunbirds; data given in Figure 5.



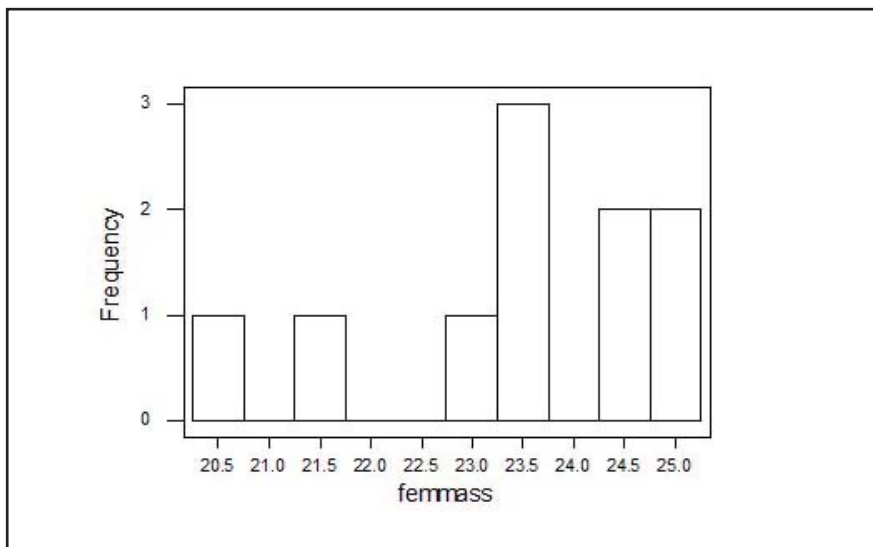


Figure 9
Mass data for female sunbirds. This histogram displays the same data as in Figure 8, but the number of intervals has been set to 10

Look at your histograms to see whether the distributions are unimodal and symmetrical. Do you think your data come from normally distributed populations?

2.1.2. Looking for relationships: Graph>Plot

This is the most useful way of displaying two variables which you think may be related in some way. In this example (Figure 10), there seems to be no relationship between the mass data for the male and female of each nesting pair, though we would probably want to see more data points to investigate this fully. In some cases, it is essential to check that you have the

dependent variable on the y-axis and the independent on the x-axis, although in our example, we could have presented the data either way round.

You can alter the formatting on most Minitab graphs by double clicking on an item on the graph. This will bring up a toolbar which allows you to insert text, or change the size or colour of a marker. You can copy and paste the finished graph into a Word document by first making sure you are in View mode (use the **Editor>View** command, then choosing to **Edit>Copy Graph**).

Graphs

The graphs that Minitab allows you to produce tend to be 'less presentable' than those in Excel and we recommend that you use Excel (where possible; for example, cluster analysis graphs can be produced in Minitab but not Excel) for the graphs that you include in your written TBA project report.

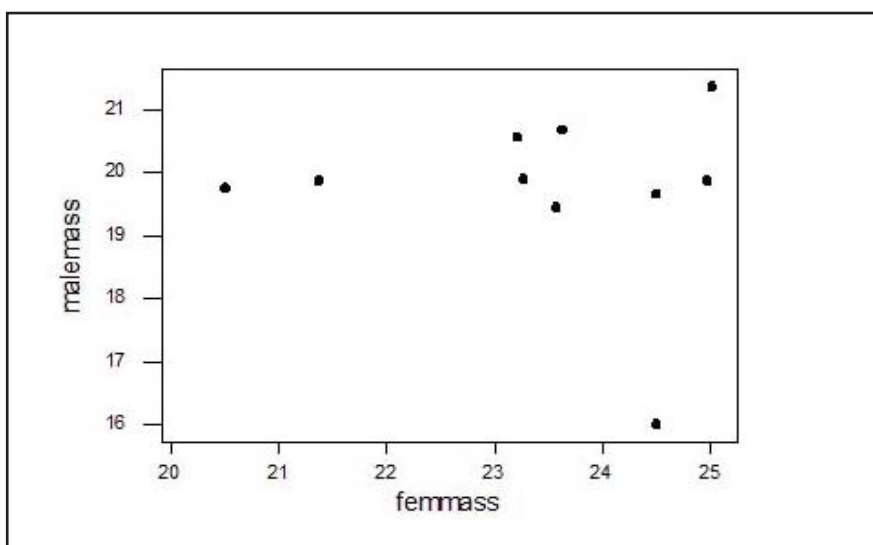


Figure 10
The relationship between the mass of males and females in 10 pairs of mated sunbirds.

Something you may wish to generate at this point is a simple output (Output 1) that gives you the descriptive statistics for your dataset (below are the 'malemass' data from Figure 5); these are used to describe the basic features of the data in a study.

They provide summaries about the sample and the measures and, together with graphics as above, they form the basis of virtually every quantitative analysis of data. For definitions of the various descriptive statistics see the TBA 'Simple guide to statistics'.

Output 1 Descriptive Statistics: malemass

Variable	N	Mean	Median	TrMean	StDev	SE Mean
malemass	10	19.624	19.781	19.748	2.396	0.758

Variable	Minimum	Maximum	Q1	Q3
malemass	15.178	23.077	17.765	21.707

'Mean mass for these males was 19.6 ± 0.8g (S.E.), n = 10'

STATISTICAL TESTING

3.1 TESTING ASSUMPTIONS

Once you have learnt to carry out and interpret the results of your statistical tests, it is important that you should be able to report the results properly. Whenever an example of the results from Minitab is presented in this guide, you will also find an example of how these results could be reported in a write-up or paper. These sections will be shown in handwriting font, to distinguish them from other text in this guide (please don't use this font in your own write-ups!!!). Obviously, you will structure exactly what you write in your reports according to the information you want to get across. But in general, you should always include essentials as shown in the tip box below.

3.1.1 Testing for normality

One of the first questions you must ask yourself before you can choose the appropriate test for your data, is whether the data set appears to be drawn from a normal distribution. You can get some idea of this by looking at a histogram of the data (section 2.1), but we generally require a more rigorous test. Minitab can do a test for normality in a number of ways, but the simplest is to choose **Stat>Basic Statistics>Normality test**. Enter the variable to be tested (Figure 11).

There is no need to enter reference probabilities. Leave the test on the default option, which is the Anderson-Darling test. Click **OK**. The output of the test is in the form of a graph (Figure 12), which shows you a plot of your data values (on the x-axis) against the cumulative probability (y-axis). The straight line is where we would expect the points to fall if the data were drawn from a normal distribution. The further the points are from a straight line, the less likely it is that your data are drawn from a normally distributed population. The null hypothesis for this test is that the distribution of your data does not differ significantly from a normal distribution, so Minitab is testing to see how likely this is. The test statistic, A^2 , and sample size, n , are given at the bottom of the plot, right hand side. The most important part for you is the p value. This gives the probability that the data come from a normally distributed population, so if p is large ($p > 0.05$ is our cut off) then we can say that the distribution is most likely normal. If p is very small ($p < 0.05$), then it is unlikely that the data are normally distributed.

'The data for mass were normally distributed; Anderson-Darling test for normality, $A^2 = 0.178$, $n = 10$, $p = 0.891$.'

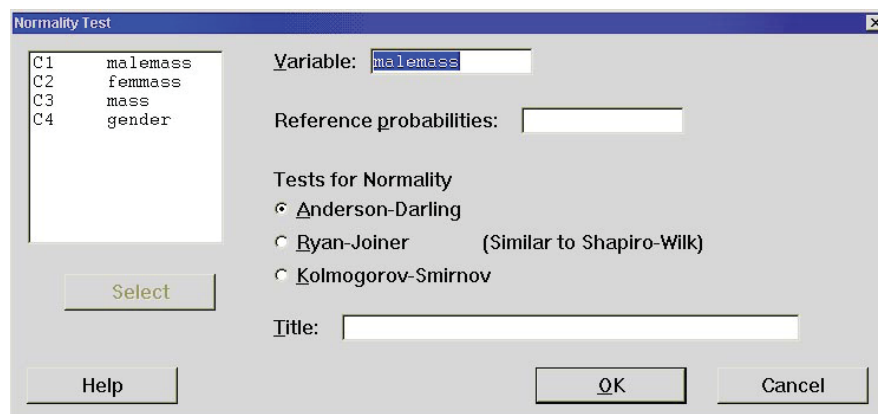
As always, report test statistic, sample size and p value.

TIP BOX - What to include when presenting your statistics

- * **Descriptive statistics** (a measure of central tendency and dispersion, and sample size) for the samples you are comparing (remember to include UNITS)
- * The **name of the test** you carried out,
- * The **test statistic**, for example, the value of t in a t -test, or F in ANOVA
- * The **degrees of freedom** and/or **n** (the number of observations in your sample)
- * The **p value**. If possible, give an exact p value; some authors just report whether $p < 0.05$ or $p > 0.05$.
How to present P values:
 $p < 0.05$ means p (probability) is less than 0.05
 $p > 0.05$ means p is greater than 0.05
Remember to check that you have the $<$ and the $>$ the right way round, as they can change the whole meaning of the test result. (Reminder: ' p ' is a probability value - where $p < 0.05$, your result tells you that the chance of you finding that effect by chance alone is less than 5% or 1 in 20.)

Figure 11

Command screen to select the Anderson-Darling test for normality on the male mass data.



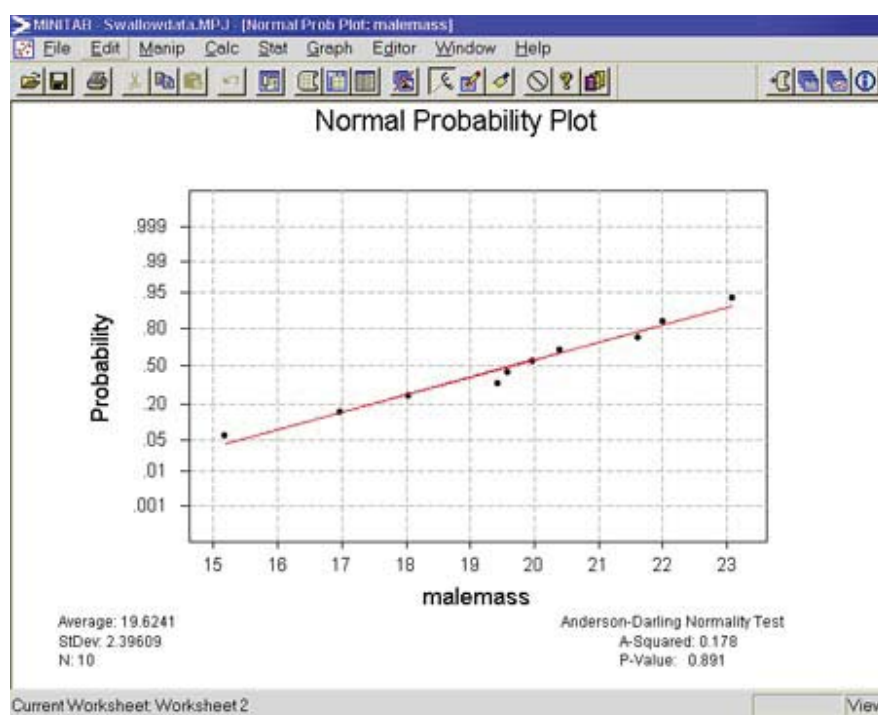
3.1.2 Homogeneity of variance

Some tests (e.g. ANOVA) require that the variance in each group of data is about the same, i.e. the variances are homogeneous. To test for this, use Minitab to perform a homogeneity of variance test. If you are comparing two variances, you can use the **Stat>Basic Statistics>2 variances** command. If you are comparing three or more variances you can find the test under **Stat>ANOVA>Test for Equal**

Variances. For the latter command you will need to have your variable stacked in a single column, with a second column indicating subscripts. Both commands conduct two tests, the F-test and Levene's test. Minitab reports a test statistic and p-value for each. Use the value for F-test unless the data are NOT normally distributed, in which case use Levene's. In either case, the null hypothesis is that the variances are equal, so if $p > 0.05$, you can assume your variances are roughly the same.

Figure 12

Normal probability plot graph for male mass using Anderson-Darling test.



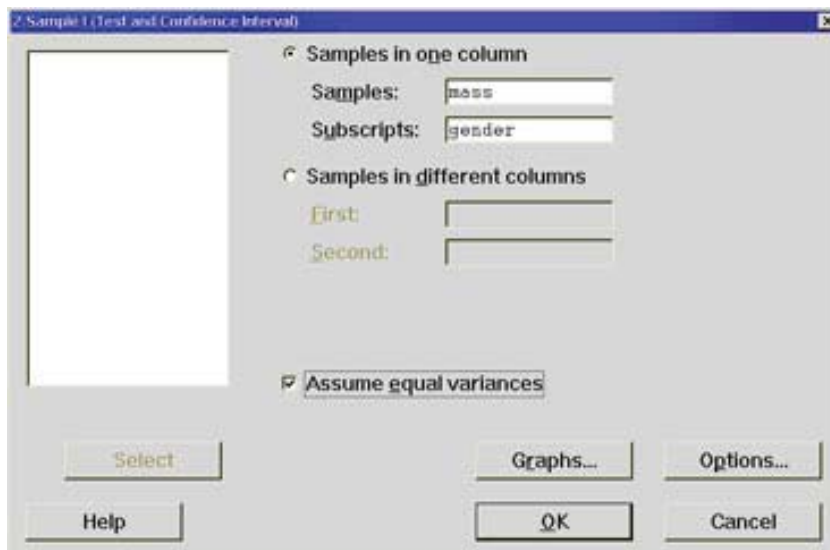
3.2 PARAMETRIC TESTS

3.2.1 Independent samples t-test

To run this test, choose **Stat>Basic Stats>2-sample t**. You can use this test with data stored in one of two ways. If your data from the two groups are listed in separate columns (unstacked), then choose the **samples in different columns** option. If the data are 'stacked' into a single column, with a second column

indicating which group each data value belongs to (see section 1.5.3), then choose **samples in one column**. In the example previously (Figure 13), we used the stacked data for male and female mass (c3) from Figure 5.

Before conducting this test you should follow instructions in section 3.1.2 to test whether the variances of the two samples are the about the same. If they are, you can

**Figure 13**

Command screen to conduct a 2-sample (or independent samples) t-test on mass, stacked into a single column (mass) with subscripts indicating gender (data in Figure 5).

check the **Assume equal variances** box. (You could try this test with and without the assumption of equal variances and note what the effects on the p value and the degrees of freedom are.)

The output of the test (Output 2 below), provides descriptive statistics for the two groups, and we can see that the mean mass for females is higher than that for males. The next line gives an interval in which we are 95% confident that the true difference between mean male and female mass will lie (remember that if the true means for male and female mass are the same, then this value will be 0). We need to concentrate on the last line. This tells us (in note form) the results of

the test of the null hypothesis that 'mu malemass' (μ or the mean mass for the male population) is equal to 'mu femmass' (the mean mass for the female population), versus the alternative hypothesis that they are not equal. The calculated t statistic in this case is $T = -5.73$, with 18 degrees of freedom. The p value ' $p = 0.000$ ' shows how likely it is to get a t statistic at least as big as this by sampling error, if the null hypothesis is really true (i.e. there really is no difference between mean male mass and mean female mass for swallows).

You can ignore the sign on the t statistic, as we could just have easily got a positive or a negative value, if the samples had been compared the other way round.

Output 2 Two-Sample T-Test and CI: mass,gender

Two-sample T for mass

gender	N	Mean	StDev	SE Mean
1	10	19.71	1.43	0.45
2	10	23.46	1.50	0.47

Difference = $\mu(1) - \mu(2)$

Estimate for difference: -3.748

95% CI for difference: (-5.122, -2.373)

T-Test of difference = 0 (vs not =): T-Value = -5.73 P-Value = 0.000 DF = 18 Both use Pooled StDev = 1.46

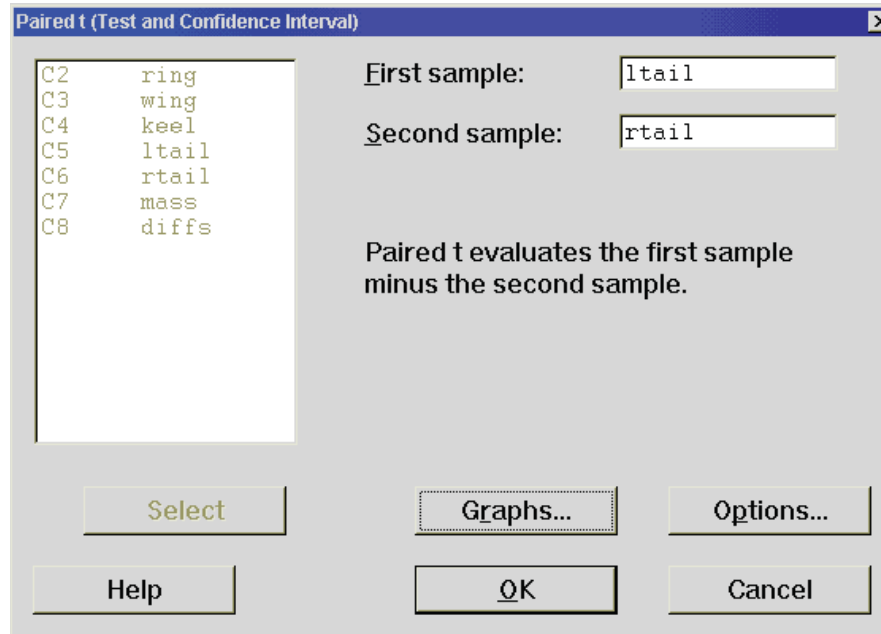
'Males were significantly lighter than females; mean male mass = $19.7 \pm 0.5g$ (mean \pm S.E.), $n = 10$, mean female mass = $23.5 \pm 0.5g$, $n = 10$, independent samples t-test, $t = 5.73$, d.f. = 18, $p < 0.001$.' See tip box p. 16.

TIP

- ✱ If you find Minitab reporting a probability of $p = 0.000$, beware. It is actually impossible for p to be equal to 0. What Minitab means is that p is smaller than 0.001, which is the smallest value Minitab is prepared to print. We should report such a result as $p < 0.001$.

Figure 14

The command screen for a paired t-test to compare the mean length of left and right tail feathers.

**3.2.2 Matched pairs (or paired samples) t-test**

If your data are paired in some way (before and after data, or data from twins for example), then the paired sample t-test is appropriate. To do this in Minitab, your data must be stored in two columns, one for each group, with the members of each matched pair in the same row. For our example, we will use the left and right tail feather lengths measured on our 10 swallows. The data are stored in columns c5 and c6 of Figure 2 (page 7).

The simplest way to conduct this test is as follows: choose **Stat>Basic Statistics>Paired t** and enter the appropriate two variables into the boxes for first and second columns as shown above in Figure 14.

This test will generate an output as shown below (Output 3), from which we can obtain the descriptive statistics for the tail lengths on either side, and conclude whether the two differ significantly (in this case they do not, as $p > 0.05$).

Alternative method. There is a more long-winded way to do this test, which nonetheless has certain advantages, as it allows you to test the assumptions of this test more easily. Recall that the paired-samples t-test requires that the differences between the paired variables be normally distributed, not that the two variables themselves are normally distributed. To test this assumption, we would need to compute a new column of differences, to which we could apply the Anderson-Darling test for normality.

So the alternative way to perform the test is first to compute a third variable, which represents the difference between the values of the paired variables for each case. Use **Calc>Calculator** to produce this new variable. You should provide a blank column for the new variable to be stored in, or enter a name for the new column (for example 'diffs') in the box to **store result in variable** and Minitab will enter the new data in the first empty column. Now type '**c5-c6**' (or wherever your data are stored) in the **expressions** box. Click OK and check the worksheet; there should be a new column containing the differences between the tail feather lengths for each bird. You can run a normality test on this column to check whether the assumption for the paired-samples t-test is fulfilled. If it is, then you now need to proceed to test the null hypothesis that there is no real difference between the means of the two variables. We do this by testing that the mean of the differences between the pairs of variables (the data stored in 'diffs') is equal to 0. So choose **Stat>Basic Statistics>1-sample t...** and enter the variable 'diffs' into the **Variables** box and fill in the box to **Test mean = 0**. The results for the example analysed in this way are displayed in the Session window (Output 4). We are interested in are the test statistic t and the p value. The degrees of freedom for this test are equal to the number of pairs (N) minus 1. Note that these values should be the same whether you conducted the test in this way, or using the paired-samples command (as above). If our value of $p >$ (is greater than) 0.05, conclude no difference between means; if $p < 0.05$, conclude that there is a difference.

Output 3 Paired T-test and CI: ltail, rtail

```

Paired T for ltail - rtail
      N      Mean      StDev      SE Mean
ltail   10    91.8400    4.8477    1.5330
rtail   10    91.4800    5.1031    1.6137
Difference  10    0.360000    1.851846    0.585605
95% CI for mean difference: (0.964731, 1.684731)
T-Test of mean difference: 0 (vs not = 0): T-value = 0.61 P-Value = 0.554

```

Output 4 One-Sample T: diffs

```

Test of mu = 0 vs not = 0
      Variable  N      Mean      StDev      SE Mean      95% CI      T      P
diffs      10    0.360000    1.851846    0.585605    (-09.464731, 1.684731) 0.61    0.554

```

'There was no significant difference between mean length for left and right tail feathers; matched pairs t-test, t = 0.61, n = 10, p = 0.554'

3.2.3 Analysis of variance**3.2.3.1 One-way ANOVA**

If you wish to perform an analysis of variance with one factor, then the **Stat>ANOVA>One-way** command is for you. Whether you choose the **One-way (unstacked)** option or not depends on how you have stored the data in the worksheet. If all the data are in a single column, with subscripts showing which group they belong to in a second column, then choose **One-way**.

If, on the other hand, the data from each group are in different columns, then choose **One-way (unstacked)**.

Although the first option is perhaps less obvious initially, it leaves you able to go on to complex analysis more easily, so if possible, enter the data in this stacked format. For our example, we will use the **One-way** command. Our data represent mass of bats caught in each of four different roost sites. We wish to know whether there is a difference in mass according to roost site.

In Figure 15, the mass data are given in column 1. In column 2, a code indicates which of the three roost sites each data point was collected from.

Choose **Statistics>ANOVA>One-way** and select the variable that you are testing (e.g. mass in c1 in our example) to go into the **Response** box. Put the column with subscripts (e.g. site in c2) in the **Factor** box. (As seen in Figure 15).

Output 5 shows the results of the test. The table shows the sources of variation, the degrees of freedom (DF),

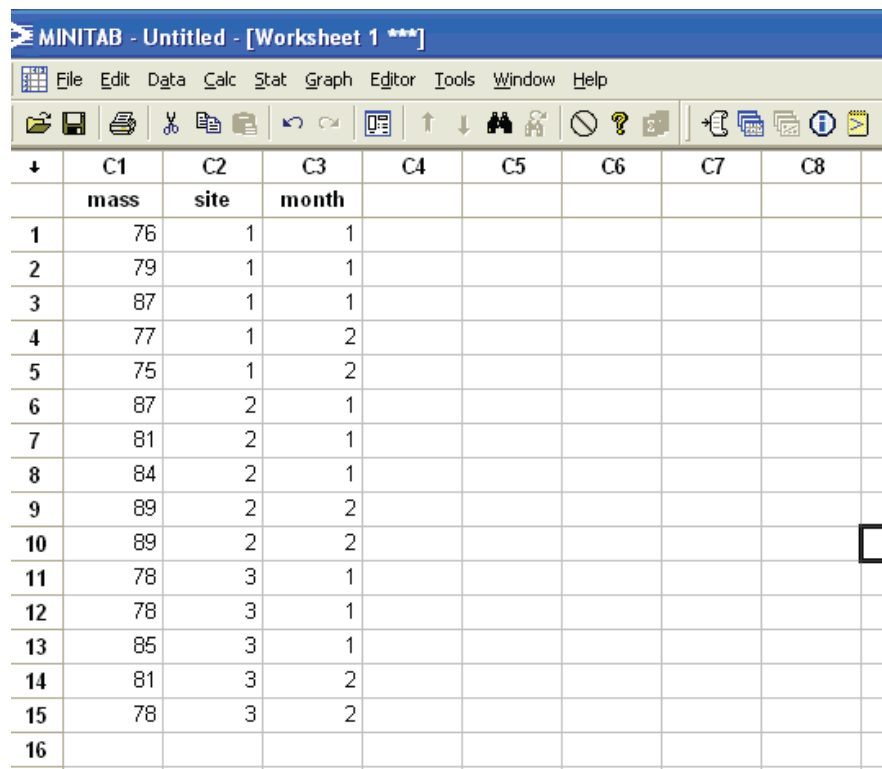
sums of squares (SS) and mean square (MS) for each variance component. The variance ratio is given by F (this is the test statistic) and, as usual, the significance of the test is shown by the p value. As always, if $p > 0.05$, there is no significant difference between groups; if $p < 0.05$, we conclude that there is a significant difference, though we have not tested to see which group(s) differ significantly from the others.

Post hoc tests

If you do have a significant difference, *post hoc* tests such as Tukey's or Fisher's are available to help you determine which groups differ from the rest. These tests make pairwise comparisons between all pairs of means. In Output 5, the small p value for the ANOVA ($p = 0.026$) indicates that the means of the groups do differ from each other. The graph which Minitab sketches for you shows the mean of each group (marked with a *) and the interval within which we are 95% confident that the true population mean lies, shown by dashed lines. From the graph in Output 5, we can see that, although the 95% confidence intervals (CI) do overlap, birds from site 2 appear to be heavier than those from the other two sites. In order to investigate this further, you might wish to employ a *post hoc* test. To do this, run the one-way ANOVA again, this time checking the box for **Comparisons** and selecting which of the tests you wish to use. The second part of Output 5 shows the results of Tukey's test on the bat mass data. The lower part of the output box shows upper and lower limits for the confidence interval for the differences between pairs of means, together with a centre value. For instance, the 95% confidence interval for the difference between

Figure 15

Worksheet showing
starling mass data
(c1) for bats from 3
different sites (coded
1-3 in column c2)
and recorded in 2
different months
(coded 1 and 2 in
column c3).



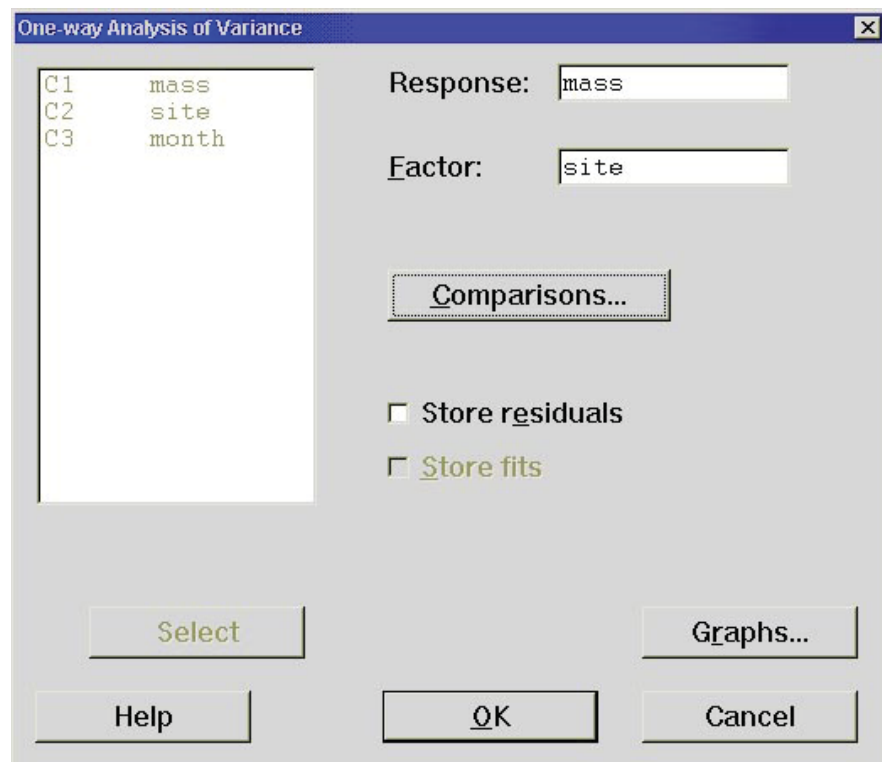
	C1	C2	C3	C4	C5	C6	C7	C8
	mass	site	month					
1	76	1	1					
2	79	1	1					
3	87	1	1					
4	77	1	2					
5	75	1	2					
6	87	2	1					
7	81	2	1					
8	84	2	1					
9	89	2	2					
10	89	2	2					
11	78	3	1					
12	78	3	1					
13	85	3	1					
14	81	3	2					
15	78	3	2					
16								

mean mass for sites 1 and 3 is between -5.31 and 7.71g. As the values encompass 0, we cannot conclude that the mean mass for these two sites differs significantly. Look for any intervals which do not contain 0 (i.e. either

all values are negative or positive). Any such intervals would indicate a significant difference between the pair of means being compared. In our example, only sites 1 and 2 seem to differ significantly.

Figure 16

Command screen to
perform a one-way
analysis of variance
to determine
whether there is a
difference between
mean mass of bats
in the three roost
sites.



One-way Analysis of Variance

Response: mass

Factor: site

Comparisons...

☐ Store residuals

☐ Store fits

Select

Help

OK

Graphs...

Cancel

Output 5 One-way ANOVA: mass versus sit

Source	DF	SS	MS	F	P
site	2	148.8	74.4	4.99	0.026
Error	12	178.8	14.9		
Total	14	327.6			

S = 3.860 R-Sq = 45.42% R-Sq (adj) = 36.32%

Individual 95% CIs For Mean Based on Pooled St Dev

Level	N	Mean	StDev
1	5	78.800	4.817
2	5	86.000	3.464
3	5	80.000	3.082

76.0 80.0 84.0 88.0

Pooled StDev = 3.860

Tukey 95% Simultaneous Confidence Intervals

All Pairwise Comparisons among Levels of site

Individual confidence level = 97.94%

site = 1 subtracted from:

site	Lower	Center	Upper
2	0.692	7.200	13.708
3	-5.308	1.200	7.708

-7.0 0.0 7.0 14.0

site = 2 subtracted from:

site	Lower	Center	Upper
3	-12.508	-6.000	0.508

'Bat mass (g) differed between roost sites; one-way ANOVA $F_{2,12} = 4.99$, $p = 0.026$.'

The numbers in subscript after the F value are the degrees of freedom for the factor (site), then the error variance components.

3.2.3.2 Two-way ANOVA

If you have two factors in your analysis of variance, the way to perform your analysis in Minitab is determined by whether your design is balanced or not (i.e. whether you have equal numbers of observations in each level of each factor) and whether your factors are fixed or random. If the design is balanced, then you should choose **Stat>ANOVA>Two-way** for fixed models or **Stat>ANOVA>Balanced ANOVA** if you have random factors. The distinction between fixed and random factors is a complicated issue, and the best advice here is to consult a statistician.

If you do not have equal numbers of observations in each cell, then you must choose **Stat>ANOVA>General Linear Model**. In either case, don't forget to enter both factors plus a third factor, the interaction term, as follows. For an example, we will use the bat mass

data of Figure 15, introducing a second factor.

It turns out that the study was also designed to investigate mass changes between two times of year, so half of the data were collected in November and half in January. We introduce 'month' as a time factor, and ask whether there is any difference in mass between the four sites or between the two collection periods. Subscripts for this factor should be listed in a third column.

We will use the General Linear Model command as we do not have equal numbers in each 'cell'. Choose **Stat>ANOVA>General Linear Model**. In the box labelled **model**, select the first factor (let's say 'site') and second factor (say 'month') from the variables list, leave a space then type 'site * month' to represent the interaction term. Your output should resemble output 6 on the next page.

Output 6 General Linear Model: mass versus site, month

Factor	Type	Levels	Values
site	fixed	3	1, 2, 3
month	fixed	2	1, 2

Analysis of Variance for mass, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
site	2	148.80	180.07	90.03	6.65	0.017
month	1	0.10	0.10	0.10	0.01	0.933
site*month	2	56.87	56.87	28.43	2.10	0.178
Error	9	121.83	121.83	13.54		
Total	14	327.60				

S = 3.67927 R-Sq = 62.61% R-Sq [adj] = 42.15%

Usual Observations for mass						
Obs	mass	Fit	SE Fit	Residual	St Resid	
3	87.0000	80.6667	2.1242	6.3333	2.11	R

R denotes an observation with a large standardised residual.

Again, you should recognise the basic format of an ANOVA table in the middle part of the output. The significance of each of the two factors and of the interaction term is shown by the p values. Always look first at the significance of the interaction term. See a text book for an interpretation of a significant interaction term. It can be helpful to plot a graph showing mean

values and confidence limits for each group to help with interpretation of results, try **Stat>ANOVA>interactions plot** or sketch the means by hand. In our example, the interaction is not significant. In such a case, it may make sense to repeat the analysis without the interaction term, to assess the significance of the two main effects (see Output 7).

Output 7 General Linear Model: mass versus site, month

Factor	Type	Levels	Values
site	fixed	3	1, 2, 3
month	fixed	2	1, 2

Analysis of Variance for mass, using Adjusted SS for Tests						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
site	2	148.80	148.80	74.40	4.58	0.036
month	1	0.10	0.10	0.10	0.01	0.939
Error	11	178.70	178.70	16.25		
Total	14	327.60				

S = 4.03057 R-Sq = 45.45% R-Sq (adj) = 30.57%

Unusual Observations for mass						
Obs	mass	Fit	SE Fit	Residual	St Resid	
3	87.0000	78.8667	1.9928	8.1333	2.32	R

R denotes an observation with a large standardised residual.

'In a two factor analysis of variance for bat mass, there was no significant interaction between the month of data collection and the roost site where birds were captured; interaction term $F_{2,9} = 2.10$, $p = 0.178$. When the analysis was repeated excluding the interaction term, mass was found to differ significantly between roost sites ($F_{2,11} = 4.58$, $p = 0.036$) but not between collection periods ($F_{1,11} = 0.01$, $p = 0.939$).'

Output 8 Correlations: wing, keel

Pearson correlation of wing and keel = 0.254
P-Value = 0.479

'In this sample, wing length did not correlate with keel length; Pearson product-moment correlation, $r = -0.254$, $n = 10$, $p = 0.479$ '

In this case, it is important to report the sign of r , as it indicates the direction of any correlation.

3.2.4 Pearson product-moment correlation

Remember that it is always worth looking at a plot of one variable against the other first, to check whether the association looks like it might be linear, or whether there is anything unusual e.g. a curve. If you suspect there may be a linear association, go on to the test.

The two variables to be correlated should be stored in separate columns. Choose **Stat>Basic Statistics>Correlation** and enter the desired variables in the box. You should choose to **Display p values**. Minitab prints the correlation coefficient (r), and this tells us something about the strength (magnitude of r) and direction (+ or -) of the correlation. The p value tells us about the significance of the correlation, if $p < 0.05$ then we conclude that the correlation is significant. The degrees of freedom for this test are equal to the number of pairs of data minus 2 ($n-2$).

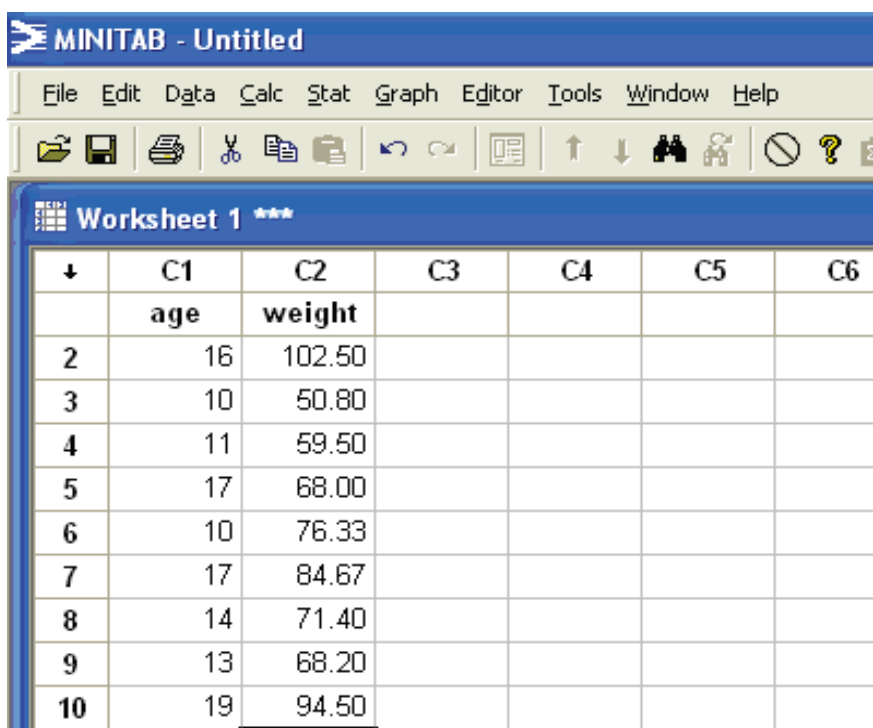
For an example, we go back to the swallow biometric data in Figure 2 (page 7), and ask whether keel length is correlated with wing length. We might expect one

to increase as the other does. Output 8 shows the correlation coefficient computed between wing length and keel length. As the p -value is greater than 0.05, we must conclude that there is no evidence for correlation between wing length and keel length in our sample.

3.2.5 Linear regression

When performing a linear regression, it is essential that you can identify which variable is dependent and which is independent (i.e. which will go on the y and x axis respectively). The two variables should be stored in two separate columns. In our example (Figure 17), we have data for the age of banded mongoose (in days after birth) in each of ten dens, and the mean weight (in g) for each litter. Weight is thought to increase linearly between 2 and 20 days since birth.

Choose **Stat>Regression>Regression**. The y variable (dependent, weight in our example) is entered in the **response** box, whilst the x (independent, age in days) variable goes in as a **predictor**. Click **OK** to run the regression.



The screenshot shows the Minitab 'Untitled' window. The menu bar includes File, Edit, Data, Calc, Stat, Graph, Editor, Tools, Window, and Help. The toolbar contains icons for file operations, editing, and statistical analysis. The 'Worksheet 1' is displayed with the following data:

	C1	C2	C3	C4	C5	C6
	age	weight				
2	16	102.50				
3	10	50.80				
4	11	59.50				
5	17	68.00				
6	10	76.33				
7	17	84.67				
8	14	71.40				
9	13	68.20				
10	19	94.50				

Figure 17
Weight (g, average for the litter) and age (number of days since birth) for 10 broods of banded mongoose

The output for this test is shown below in Output 9. The first line gives an essential piece of information: the equation of the line of best fit. Remember that this is the equation of a straight line: it is of the general form ' $Y = mX + c$ ', (except that Minitab writes it as $Y = c + mX$)

where m is the gradient or slope, and c is a constant. The line will cross the y -axis at the co-ordinates $(0, c)$. In our example below, the equation of the regression line (weight = $9.2 + 4.52 \text{ age}$) tells us that there is an increase in weight of 4.52g for every day increase in age.

Output 9 Regression Analysis: weight versus age

The regression equation is
weight = 9.2 + 4.52 age

Predictor	Coef	SE Coef	T	P
Constant	9.22	18.80	0.49	0.637
age	4.520	1.342	3.37	0.010

S = 14.2290 R-Sq = 58.6% R-Sq (adj) = 53.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2296.4	2296.4	11.34	0.010
Residual Error	8	1619.7	202.5		
Total	9	3916.1			

EXAMPLE -

The next part of the output provides a t -test of two hypotheses. The first is a test of the null hypotheses that the constant is really no different from 0. The p value for this test ($p = 0.637$) is greater than our critical value of 0.05, so we conclude that the constant in the equation is not significantly different from 0. In our example, the meaning of this is somewhat confused, by the fact that we wouldn't expect the equation to be linear for mongoose younger than 2 days (0 day old mongoose haven't been born yet, i.e. difficult to weigh!)

The second of these tests is arguably the most important. It is labelled 'age' and it tests the null hypothesis that the slope of the line is really no different from 0. If the slope is not different from 0, we would really have no relationship between the two variables, and the regression would be meaningless. In our case, $p = 0.010$ so we can conclude, in this case, that the slope is significantly different from 0, therefore the regression is meaningful. The degrees of freedom for this test are $(n-2)$.

Our other main point of interest in the output is the value of **R-sq** (R^2), or the coefficient of determination. Use the adjusted value **R-sq(adj)** as this has been modified to account for the degrees of freedom. Remember that R^2 gives a measure of how much of the variation in y is statistically accounted for by the variation in x . In our example, 54% of the variation is accounted for by the regression.

The analysis of variance test essentially performs the same function as the t -test for the significance of the slope. The table shows a p -value, which will always be equal to the p value given for the t -test of the slope above. You will often see the F statistic quoted to show the significance of the regression.

Minitab also lists any unusual values; these are marked with an X if it is the x value which is unusual or an R if it is the response (or y value) which is unusual. These may be errors in the data, or could be points which show

something interesting, by being very different from the rest. In either case, they are worth investigating:

'For mongoose aged between 2 and 20 days, weight (g) is linearly related to age (days after birth), according to the equation 'weight = $9.2 + 4.52 \text{ age}$ ', $r^2 = 53.5\%$, $F_{1,8} = 11.34$, $p = 0.010$.'

TIP BOX - What to include when presenting your statistics

✱ Another useful command in Minitab is the **Stat>Regression>Fitted Line Plot**, which plots your data and the fitted line together. You can add 95% confidence bands to the plot, if you check that option.

3.3 NON-PARAMETRIC TESTS

These are the tests to use if your data do not fit a normal distribution, or if you have too few data to tell, with any confidence, whether the distribution is normal or not. Non-interval data also require non-parametric tests.

3.3.1. Mann-Whitney test

This test is the non-parametric equivalent of the independent samples t-test; it looks for a difference in the distributions of the two samples. You don't need to have the same number of data values in the two samples.

You should have the two variables stored in two separate columns. Choose **Stat>Nonparametrics>Mann-**

Whitney, and put the first and second samples into their respective boxes. Leave the confidence level at 95% and the alternative hypothesis that the medians are **not equal**. Click **OK** to run the test.

A typical output is shown in Output 10. The two data sets (not shown) recorded the number of 'fruits' on two trees (labelled tree 1 and tree 2) of *Ficus*. The medians of the two are the same in this case. The test statistics Minitab has calculated is the W statistic (hand calculations will usually give a U value). To assess the significance of the test, use the p value which is adjusted for ties.

Output 10 Mann-Whitney Test and CI: tree 1, tree 2

		N	Median
tree	1	9	12338
tree	2	7	12338

Point estimate for ETA1-ETA2 is 1112
 95.6 Percent CI for ETA1-ETA2 is (-3805, 8150)
 W = 81.5
 Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.6338
 The test is significant as 0.6336 (adjusted for ties)

Cannot reject at alpha = 0.05

'There was no difference between the number of fruits in tree 1 (median = 12338 fruits per tree, n = 9) and tree 2 (median = 12338 fruits per tree, n = 7); Mann-Whitney test, W = 81.5, p = 0.634.'

For a full report, you should also calculate the IQR for the two variables, to include a measure of dispersion for each set of trees.

EXAMPLE -

- * Minitab reports the results of this test as 'significant at 0.6336'. DON'T BE FOOLED. You know that differences are only significant if p is smaller than 0.05, and 0.63 is not smaller than 0.05. This result is therefore not significant, and we must accept that there is no difference between the medians of samples from the two lines. In fact, Minitab does confirm this in the final line: 'cannot reject (the null hypothesis) at alpha = 0.05'. The results of this test are not clearly worded and many people are fooled by it... don't let it be you!
- * IQR – the inter quartile range is a measure of statistical dispersion, and is equal to the difference between the third and first quartiles. As 25% of the data are less than or equal to the first quartile and the same proportion are greater than or equal to the third quartile, the IQR will include about half of the data. The IQR has the same units as the data and, because it uses the middle 50%, is not affected by outliers or extreme values. The IQR is also equal to the length of the box in a box plot. For more detail, and information about other common examples of measures of statistical dispersion (e.g. variance, standard deviation), see the TBA 'Simple guide to statistics'.

3.3.2 Wilcoxon signed rank test

This test is the non-parametric equivalent of the matched pairs t-test. The way that it is calculated on Minitab is similar to the alternative calculation of the paired-samples t-test (section 3.2.2). Because you are

calculating differences between the matched values, some books recommend that data should be interval, (not ordinal) for the test to work, but in fact the test is commonly used with ordinal data.

You must store the data for the two variables in two separate columns, with the corresponding values for each matched pair in the same row. Use **Calc>Calculator** to produce a new variable which represents the difference between each pair of values, by typing ' $c1 - c2$ ' (or whatever your columns are) into the **expressions** box, and providing an empty column for the new variable to be stored in. You must then use the Wilcoxon test to check whether the median of these differences is significantly different from zero, as follows. Choose

Stat>Nonparametrics>1 sample Wilcoxon and enter the column of differences into the **Variables** box. Check the box marked **Test median**, and leave the value as **0.0** with the alternative hypothesis set as **not equal**. Minitab will produce a Wilcoxon statistic (rather than T produced in hand calculations) and a p value. As always, if $p > 0.05$, conclude that there is no significant difference between the distributions of the two groups; if $p < 0.05$, the two do differ significantly.

Output 11 Wilcoxon Signed Rank Test:

Test of median = 0.000000 versus median not = 0.000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diffs	10	10	1.5	0.009	-2.000

'There was a significant difference between the median number of visual displays performed in a 10 minute trial by male *P. krefftii* in the presence of familiar versus novel females; Wilcoxon signed-rank test, $W = 1.5$, $n = 10$, $p = 0.009$.'

Again, for a full report, calculate and report the median and IQR for each group.

Output 11 shows the results of an experiment to investigate the number of times in which male *Phrynobatrachus krefftii* frogs perform a visual display (inflation of their colourful throat patch) in a 10 minute trial (data not shown). Each male was observed with a familiar female and, in a separate trial, with a novel female.

3.3.3 Kruskal-Wallis test

This is the non-parametric equivalent of a one-way ANOVA, and is used to test for a difference between the medians of 3 or more groups of data. For this test, you must have your data from each group stored in the same column, with a second column providing subscripts to identify the group to which each value belongs.

Output 12 Kruskal-Wallis Test: pollinators versus Kruskal

Kruskal-Wallis Test on pollinators

Kruskal	N	Median	Ave Rank	Z
1	10	4.000	17.7	0.97
2	10	3.500	13.0	-1.10
3	10	4.000	15.8	0.12
Overall	30		15.5	

$H = 1.44$ $DF = 2$ $P = 0.486$

$H = 1.52$ $DF = 2$ $P = 0.466$ (adjusted for ties)

'There is no significant difference between the median number of visits by pollinators to flowers in each of three colour categories; median number of visits per flower = 4 for category 1, 3.5 for category 2 and 4 for category 3: Kruskal-Wallis test, $H = 1.53$, d.f. = 2, $p = 0.466$.'

Again, you should also include IQR for each group.

Choose **Stat>Nonparametrics>Kruskal-Wallis**, and, as for ANOVA, list the data variable in the **Response** box, and the column containing subscripts in the **Factor**. An example of the results of an experiment to compare the number of visits by pollinators in a one minute observation period, to flowers of three different colour categories is shown in Output 12. The table provides a summary of the number of observations in each sample (labelled NOBS !!), and the median of each sample.

We use the test statistic **H-adj**, as this is adjusted for any ties in the data. The degrees of freedom are equal to the number of groups minus 1. If $p < 0.05$, the result is significant, which means that there is a significant difference between at least two of the medians (you don't know which ones differ, but a quick look at the data can often be revealing). If $p > 0.05$, there is no significant difference between any of the group medians.

3.3.4 Spearman rank correlation

To perform this test on Minitab, you must first rank the data for each of your two variables. Choose **Data>Rank** and place the first variable to be ranked in the **Rank data in** box. Find an empty column to put the ranks in. Click **OK** then label the new column with an appropriate name, (such as 'ranks A') as it would be easy to confuse it at a later stage. Now produce ranks for the second variable in the same way, and again, label the new column with something sensible. You can now correlate the two columns of ranks (**Stat>Basic Statistics>Correlation**, following the instructions in section 3.2.4). Although the result is listed as a Pearson correlation, you have, in fact, carried out a Spearman rank correlation, as you did the test on the ranks, not the raw data.

Output 13 Correlations: ranks bee, ranks flowers

Pearson correlation of ranks bee and ranks flowers = 0.181
P-Value = 0.520

Bee species diversity (the number of species located on a 1km transect) was not correlated with floral diversity (number of species in bloom along the transect); Spearman rank correlation, $r = 0.181$, d.f. = 13, $p = 0.520$.

Output 13 is the result of a Spearman rank correlation between the number of different bee species located along a transect, with the number of flower species in bloom. Data (not shown) were available for 15 transects.

3.3.5 Chi-square test

3.3.5.1. Contingency table

If you have data for a contingency table, Minitab will produce a value for Chi-square for you. Enter the data for each group into a separate column and choose **Stat>Tables>Chi-square**. Enter the columns which contain the table. Minitab will produce a value for chi-square, and a p value (Output 14). For example, if we wanted to test whether the proportion of successful matings in a group of kobs differed between territory-holding males and 'sneaker' males, the data for the contingency table (below) would be entered as shown in Figure 18.

3.3.5.2 Goodness-of-fit test

If you need to perform a goodness-of-fit test, rather than a contingency table, you will need to enter two columns of data, one containing the 'observed' and one containing the 'expected' values. You cannot use Menu commands to calculate a value for chi-square, but

must type commands directly into the Session window instead. To do this, go to **Editor** and select **Enable Commands**. This will produce the **MTB>** prompt in the session window and allow you to type commands after the prompt. It is important to get all punctuation and spaces etc. exactly right. Go to the Session window and type:

```
MTB>let k1=sum((c1-c2)**2/c2)
```

```
MTB>print k1
```

Note that the notation ****2** means 'square the preceding expression'.

k1 is the chi-square value for the test.

To obtain an exact p value for this value of chi-square, type the following commands directly into the session window, but instead of *df* you should type the number which represents the degrees of freedom for your data, which is equal to the number of observed values minus 1 ($n-1$).

```
MTB>cdf k1 k2;
```

```
SUBC>chisquare df.
```

```
MTB>let k3=1-k2
```

```
MTB>print k3
```

Figure 18
Minitab worksheet demonstrating the layout of a data table for successful and failed mating attempts and type of male strategy for a chi-square contingency test.

	C1	C2	C3	C4	C5	C6	C7	C8
	success	failure						
1	350	90						
2	120	300						
3								
4								
5								
6								
7								
8								
9								
10								

What you are doing here is asking Minitab to calculate the cumulative distribution function (cdf) of your value of chi-square (k1) and to store it in k2. The cdf represents the probability of getting a value of chi-square less than or equal to our value. We are interested in the probability of getting a value at least as big as our value, so we calculate k3, which is equal to 1 minus k2.

The example opposite (Output 15) compares the number of different coloured fruits out of 160, in each of four colour categories that were taken by foraging chimps compared with the frequencies expected from their availability in the population.

Output 14 Chi-Square Test: success, failure

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	success	failure	Total
1	350	90	440
	240.47	199.53	
	49.895	60.129	
2	120	300	420
	229.53	190.47	
	52.270	62.993	
Total	470	390	860

Chi-Sq = 225.287, DF = 1, P-Value = 0.000

'The proportion of successful matings differed according to male strategy, chi-square = 225.29, d.f. = 1, p < 0.001*'.

* Again, be careful not to express the result as 'p=0.000'.

Output 15

```
MTB      >      let k1 = sum ((c1-c2)**2/c2)
MTB      >      print k1
Data Display
```

```
K1              1.33333
```

```
MTB      >      cdf k1 k2;
SUBC     >      chisquare 3.
MTB      >      let k3=1-k2
MTB      >      print k3
```

```
Data Display
```

```
K3              0.721233
```

'The observed fruit colour selection by chimps did not differ significantly from that expected according to availability; chi-squared goodness-of-fit test, $\chi^2 = 1.333$, d.f.= 3, p = 0.721.'

That's all folks.....but don't forget there is lots of help under the Help command in Minitab if you are stuck.

Some textbooks also give help in conducting and interpreting tests using Minitab: try Dytham, C., 1999, *Choosing and Using Statistics: a Biologist's Guide* which is available in the Library. Good luck and happy analysis!



TROPICAL BIOLOGY ASSOCIATION

Skills Series

This **Minitab guide** was developed to complement the statistics teachings on the Tropical Biology Association's field courses. These ecology and conservation field courses are based in East Africa and Madagascar. They are a tool to build capacity in tropical conservation. Lasting one month, the courses provide training in current concepts and techniques in tropical ecology and conservation as well skills needed for designing and carrying out field projects. Over a 120 conservation biologists from both Africa and Europe are trained each year.

Tropical Biology Association

The Tropical Biology Association is a non-profit organization dedicated to providing professional training to individuals and institutions involved in the conservation and management of tropical environments. The TBA works in collaboration with African institutions to develop their capacity in natural resource management through field courses, training workshops and follow-up support.

European Office

Department of Zoology
Downing Street
Cambridge CB2 3EJ
United Kingdom
Tel./Fax + 44 (0)1223 336619
email:tba@tropical-biology.org

African Office

Nature Kenya
PO BOX 44486
00100 - Nairobi, Kenya
Tel. +254 (0) 20 3749957
or 3746090
email:tba-africa@tropical-biology.org

www.tropical-biology.org