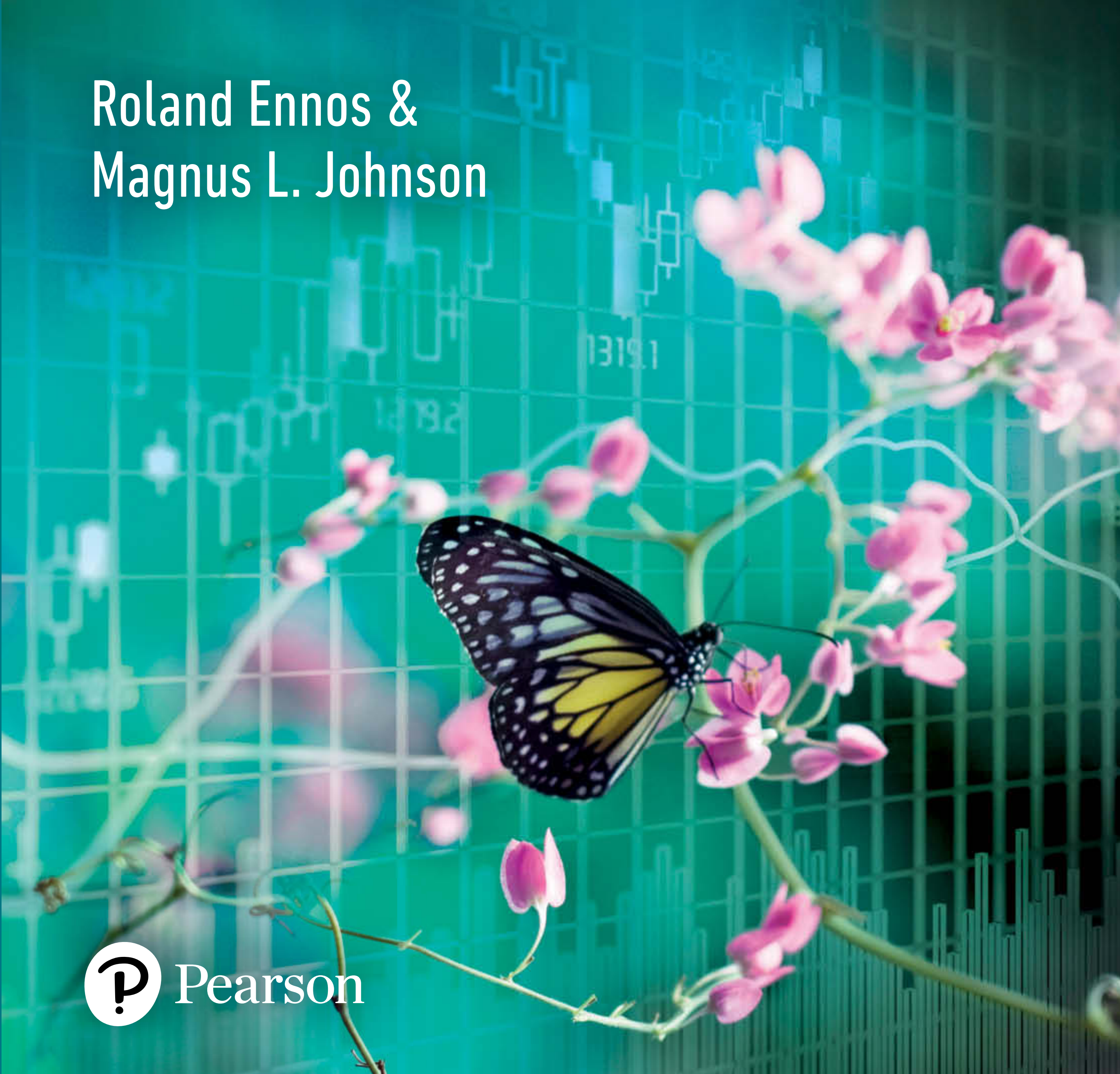
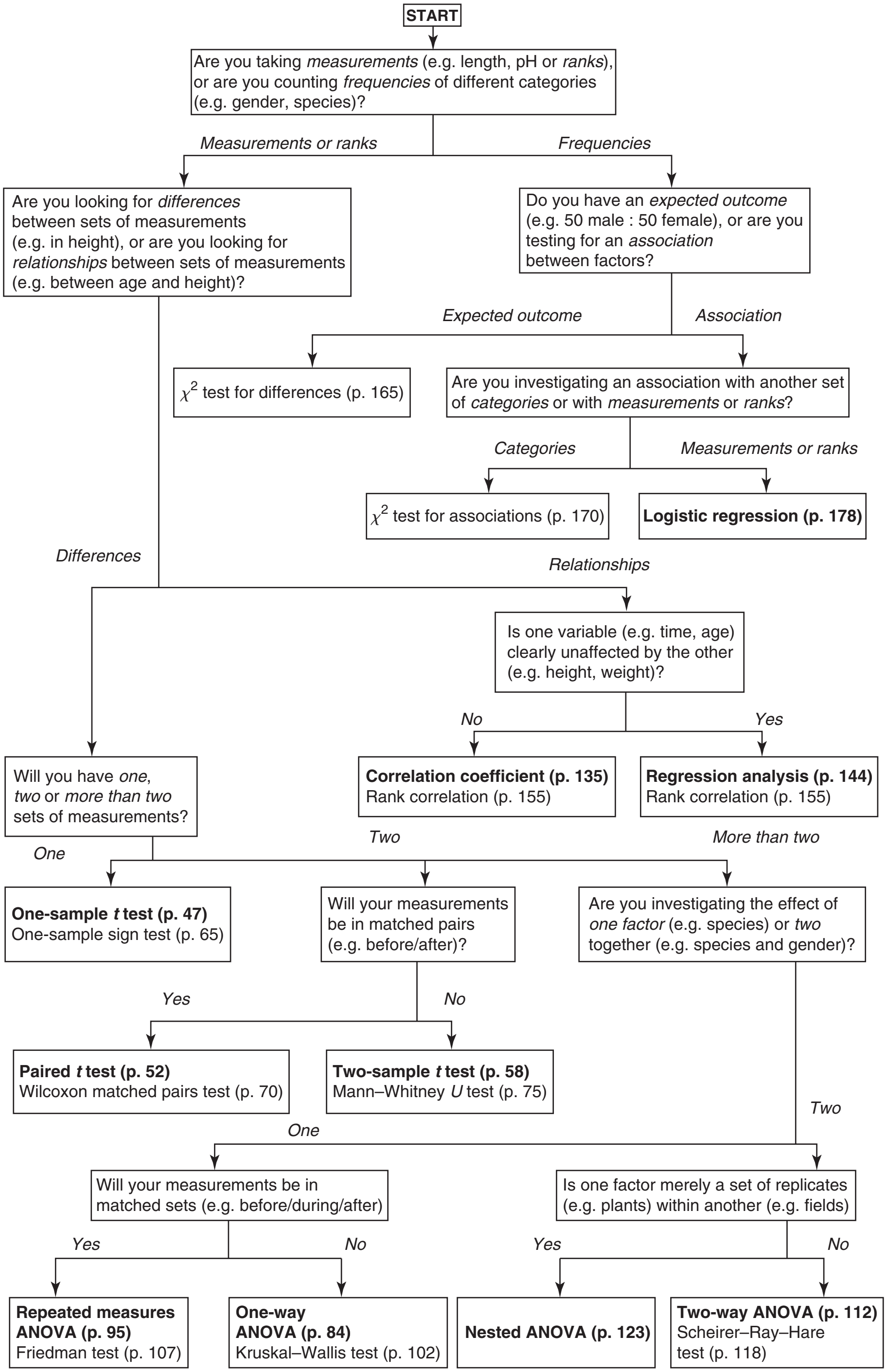
Statistical and Data Handling Skills in Biology

Fourth Edition

Roland Ennos &
Magnus L. Johnson



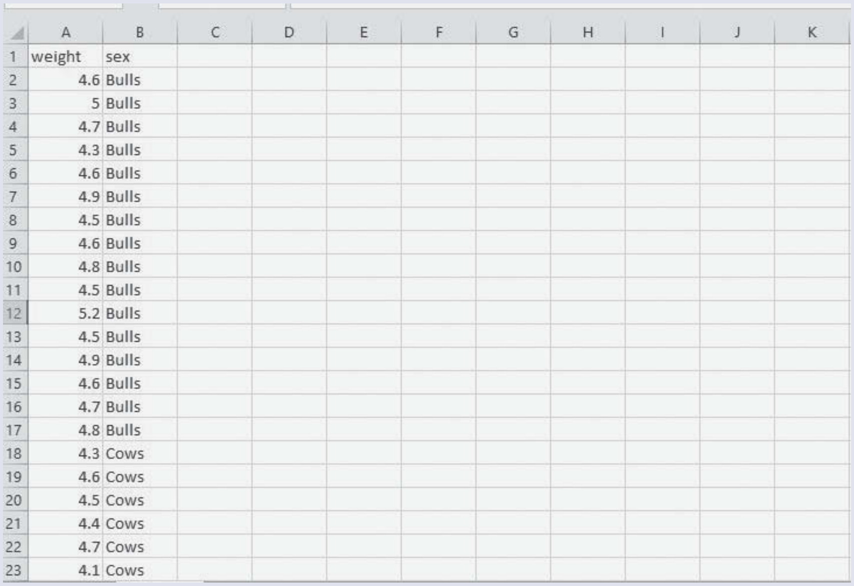
Pearson


START
Are you taking *measurements* (e.g. length, pH or *ranks*), or are you counting *frequencies* of different categories (e.g. gender, species)?
Measurements or ranks
Frequencies
Are you looking for *differences* between sets of measurements (e.g. in height), or are you looking for *relationships* between sets of measurements (e.g. between age and height)?
Do you have an *expected outcome* (e.g. 50 male : 50 female), or are you testing for an *association* between factors?
Expected outcome
Association
χ^2 test for differences (p. 165)
Are you investigating an association with another set of *categories* or with *measurements* or *ranks*?
Categories
Measurements or ranks
χ^2 test for associations (p. 170)
Logistic regression (p. 178)
Differences
Relationships
Is one variable (e.g. time, age) clearly unaffected by the other (e.g. height, weight)?
No
Yes
Correlation coefficient (p. 135) Rank correlation (p. 155)
Regression analysis (p. 144) Rank correlation (p. 155)
Will you have *one*, *two* or *more than two* sets of measurements?
One
Two
More than two
One-sample *t* test (p. 47) One-sample sign test (p. 65)
Will your measurements be in matched pairs (e.g. before/after)?
Are you investigating the effect of *one factor* (e.g. species) or *two* together (e.g. species and gender)?
Yes
No
Paired *t* test (p. 52) Wilcoxon matched pairs test (p. 70)
Two-sample *t* test (p. 58) Mann–Whitney *U* test (p. 75)
Two
One
Will your measurements be in matched sets (e.g. before/during/after)
Is one factor merely a set of replicates (e.g. plants) within another (e.g. fields)
Yes
No
Yes
No
Repeated measures ANOVA (p. 95) Friedman test (p. 107)
One-way ANOVA (p. 84) Kruskal–Wallis test (p. 102)
Nested ANOVA (p. 123)
Two-way ANOVA (p. 112) Scheirer–Ray–Hare test (p. 118)

are given at the left of the second table. In the Levene's test, if Sig. < 0.05, then the test is not valid. Here, Sig. $= 0.140$, so you could use either test. If in doubt, however, use the second, more accurate statistic.

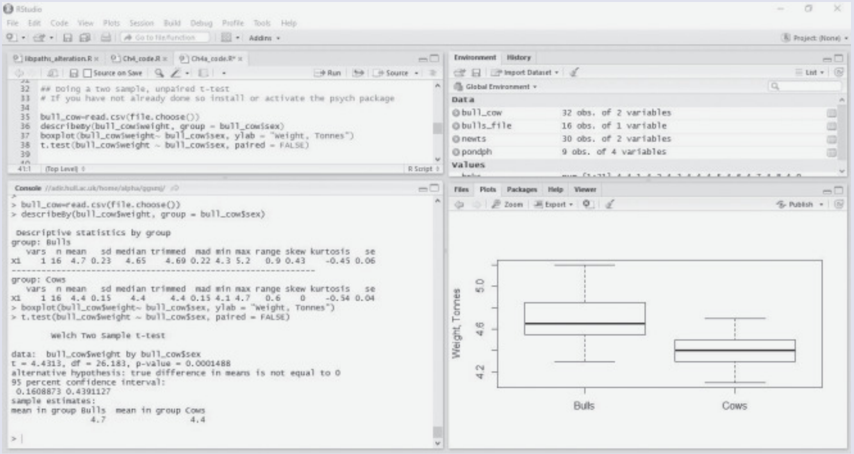
Using RStudio

There are two ways in which you can input your data to a CSV file for analysis with RStudio to perform a two-sample *t* test. You can either put the results from the two samples into separate columns or put all the data into one column (we recommend this so that later, more complex tests don't come as so much of a shock).



	A	B	C	D	E	F	G	H	I	J	K
1	weight	sex									
2	4.6	Bulls									
3	5	Bulls									
4	4.7	Bulls									
5	4.3	Bulls									
6	4.6	Bulls									
7	4.9	Bulls									
8	4.5	Bulls									
9	4.6	Bulls									
10	4.8	Bulls									
11	4.5	Bulls									
12	5.2	Bulls									
13	4.5	Bulls									
14	4.9	Bulls									
15	4.6	Bulls									
16	4.7	Bulls									
17	4.8	Bulls									
18	4.3	Cows									
19	4.6	Cows									
20	4.5	Cows									
21	4.4	Cows									
22	4.7	Cows									
23	4.1	Cows									

Call it something like *weight*. To define the two groups, you must create a second subscript (or factor) column which we will call sex. The data will be similarly arranged, therefore to the data for SPSS shown in the previous screenshot.



```
## Doing a two sample, unpaired t-test
# If you have not already done so install or activate the psych package

bull_cow=read.csv(file.choose())
describeBy(bull_cow$weight, group = bull_cow$sex)
boxplot(bull_cow$weight~ bull_cow$sex, ylab = "Weight, Tonnes")
t.test(bull_cow$weight ~ bull_cow$sex, paired = FALSE)
```

```
Data
bull_cow           32 obs. of 2 variables
bulls_file         16 obs. of 1 variable
newts              30 obs. of 2 variables
pondph             9 obs. of 4 variables
Values
```

```
> bull_cow=read.csv(file.choose())
> describeBy(bull_cow$weight, group = bull_cow$sex)

 Descriptive statistics by group
group: Bulls
   vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 16  4.7 0.23   4.65    4.69 0.22 4.3 5.2   0.9 0.43    -0.45 0.06
------------------------------------------------------------------------
group: Cows
   vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 16  4.4 0.15    4.4     4.4 0.15 4.1 4.7   0.6    0    -0.54 0.04
> boxplot(bull_cow$weight~ bull_cow$sex, ylab = "Weight, Tonnes")
> t.test(bull_cow$weight ~ bull_cow$sex, paired = FALSE)

        Welch Two Sample t-test

data:  bull_cow$weight by bull_cow$sex
t = 4.4313, df = 26.183, p-value = 0.0001488
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1608873 0.4391127
sample estimates:
mean in group Bulls  mean in group Cows
                4.7                 4.4

> |
```

To carry out the *t* test, enter the code the same as for a paired *t* test but set the command *paired* to FALSE. The commands are shown in the top left window above. From the output (bottom left), we can see that the *t* value is 4.4313 with just over 26 degrees of freedom.

Step 3: Calculating the significance probability

Using a calculator

You must compare your value of $|t|$ with the critical value of the *t* statistic for $N_A + N_B$-two degrees of freedom, where N_A and N_B are the sample sizes of groups A and B.

Here, there are $16 + 16 - 2 = 30$ degrees of freedom; the critical value of *t* for the 5% level is 2.042.

Using SPSS and RStudio

SPSS and RStudio will directly work out the probability, Sig. (two-tailed) or p-value. (Note that the bigger the value of $|t|$, the smaller the value of Sig. (two-tailed) or p-value.)

In SPSS, Sig. (two-tailed) = 0.000, and in RStudio, p-value = 0.0001488.

Step 4: Deciding whether to reject the null hypothesis

Using a calculator

- If $|t|$ is greater than or equal to the critical value, you must reject the null hypothesis. Therefore, you can say that the mean difference is significantly different from zero.
- If $|t|$ is less than the critical value, you have no evidence to reject the null hypothesis. Therefore, you can say that the mean difference is not significantly different from zero.

Here, $4.43 > 2.042$.

Using SPSS and RStudio

- If Sig. (two-tailed) or p-value ≤ 0.05, you must reject the null hypothesis. Therefore, you can say that the mean difference is significantly different from zero.
- If Sig. (two-tailed) or p-value > 0.05, you have no evidence to reject the null hypothesis. Therefore, you can say that the mean is not significantly different from zero.

Here, Sig. (two-tailed) = 0.000, p-value = $0.0001488 < 0.05$

Therefore, we must reject the null hypothesis. We can say that the mean weights of the bull and cow elephants were different. In fact, the bulls were significantly heavier than the cows.

Step 5: Calculating confidence limits

The 95% confidence intervals for the mean difference are given by the equation

$$95\%\ \text{CI(difference)} = \bar{x}_\text{A} - \bar{x}_\text{B} \pm (t_{N_A+N_B-2}(5\%) \times \overline{SE}_d) \qquad (4.7)$$

➡

Using a calculator

$$\begin{aligned} 95\%\ \mathrm{CI(difference)} &= 4.70 - 4.40 \pm (2.042 \times 0.0680) \\ &= 0.16 \text{ to } 0.44 \end{aligned}$$

Using SPSS and RStudio

SPSS and RStudio calculate these limits directly, and you should round up to two significant figures to give 0.16 to 0.44.

4.8.5 Presenting the results

If you have investigated just one or two measurements, probably the best way to present your data is in a bar chart with the means and standard errors of the two sets of data. You can then refer to this chart in the text, where you can emphasise the size of the difference and back this up with the results of the *t* test itself. For the results of the bull and cow elephants, you would present the results as follows.

The mean weights of the bull and cow elephants are shown in Figure 4.4. The bull elephants were on average around 7% heavier than the cows, a difference that a two-sample *t* test showed was significant 6.8 ($t_{26.1} = 4.43, P < 0.0005$).

Note that the p-value given is not zero as SPSS reports. Indeed, p-values are *never* zero, so when SPSS gives a value of 0.000, it just means that it is less than 0.0005. A two-sample *t* test is usually only significant if the error bars on your bar chart do not overlap each other.

If you have taken several measurements on the same samples (for instance, comparing a control and experimentally treated group), it is usually best to present the information in the form of a table as in Table 4.1, with the benefit of an informative legend.

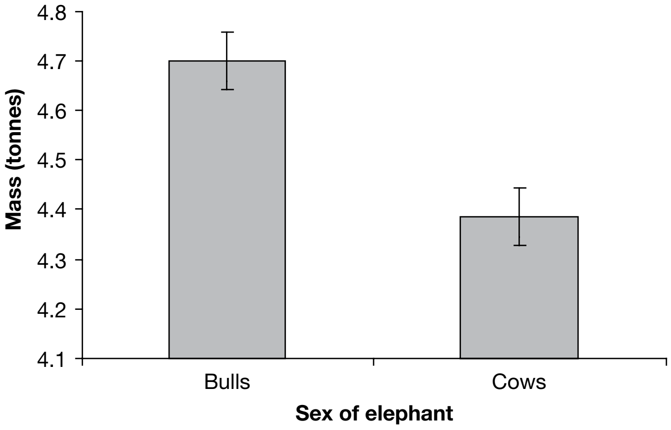


Figure 4.4 The mean (± standard error) of the masses of 16 bull and 16 cow elephants.

Table 4.1 The effect of nitrogen treatment on sunflower plants. The results show the means ± standard error for control and high nitrogen plants of their height, biomass, stem diameter and leaf area.

Treatment	*n*	Height (cm)	Biomass (g)	Stem diameter (mm)	Leaf area (cm^2)
Control	12	36.5 ± 3.6	14.3 ± 2.8	9.6 ± 1.9	194 ± 38
High *N*	12	50.4 ± 3.9**	22.45 ± 3.2***	10.9 ± 2.4 NS	286 ± 32*

Asterisks denote the degree of significance: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; NS no significant difference.

The use of asterisks to denote the degree of significance is a common method that is very concise and clear. Having produced the table, you should refer to it in the text of the results section, saying something like:

> The effects of supplementary nitrogen are summarised in Table 4.1. It can be seen that supplementary nitrogen increased height by around 38%, biomass by 57% and leaf area by 46%, but had no significant effect on stem diameter.

4.9 Introduction to non-parametric tests for differences

rank
Numerical order of a data point.

median
The central value of a distribution (or average of the middle points if the sample size is even).

Non-parametric tests don't assume that the distribution is normal and use only information about the **rank** of each data point. The tests for differences compare the **medians** of the groups instead of comparing means, and they look at the probability of getting the ranked data points in a particular order. This means that the tests are intuitively fairly easy to understand. However, it often takes a great deal of time to assign ranks and to manipulate these ranks to produce the test statistic. Therefore, it is often quicker to carry out non-parametric statistics using computer packages which do all that for you. For this reason, we will take a brief look at the rationale and mathematics of each of the non-parametric equivalents of *t* tests, before looking at how to carry it out both on a calculator and using statistical packages.

4.10 The one-sample sign test

4.10.1 Purpose

To test whether the sample median of one measurement taken on a single population is different from an expected value *E*. It is the non-parametric equivalent of the one-sample *t* test.

4.10.2 Rationale

Like the one-sample *t* test, the first stage of the one sample sign test is to calculate the difference *d* between each measurement you have made and the expected value *E*. Next, you rank the absolute values of these differences and give the positive differences a plus and the negative differences a minus sign. Finally, you add all the negative ranks together, and separately add all the positive ranks together to give two values of T: $T-$ and $T+$. Note that if the median of your sample is lower or higher than the expected median, one value of T will be large and the other will be small. In this test, the smaller T value is compared with the critical values table for the relevant group size. The null hypothesis is rejected if your smaller value of T is **lower than or equal to** a critical value. Note that this test is a special case of the Wilcoxon matched pairs test (Section 4.11) but substituting expected values for one of the two samples.

4.10.3 Carrying out the test

The method of carrying out the test is best seen by working through Example 4.4.

Example 4.4

After a course on statistics, students were required to give their verdict on the merits of the course in a questionnaire. Students could give scores of 1 = rubbish, through 3 = reasonable, to 5 = excellent. The following scores were obtained:

Number	
Students scoring 1	8
Students scoring 2	14
Students scoring 3	13
Students scoring 4	4
Students scoring 5	0

Did the course score significantly lower than 3?

Solution

Step 1: Formulating the null hypothesis

The null hypothesis is that the median of the scores the students gave was 3.

Step 2: Calculating the test statistic

Using a calculator

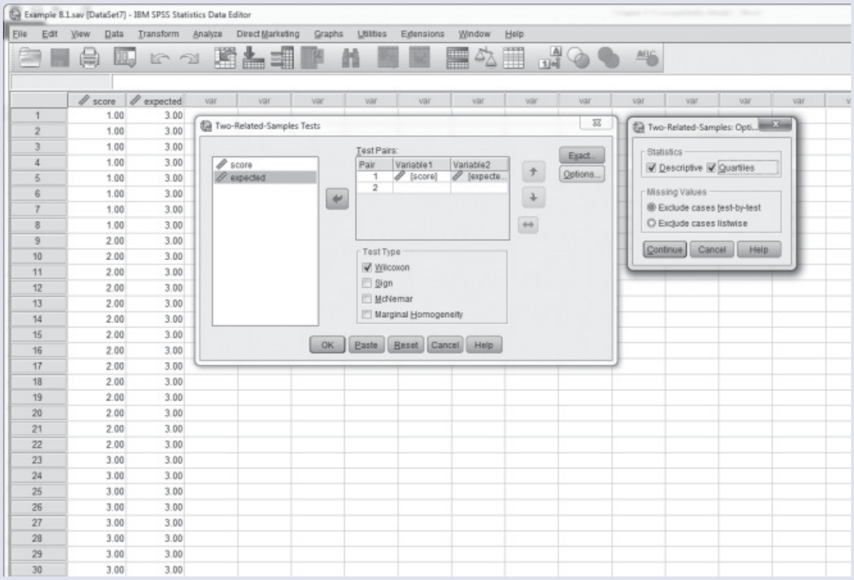
	N	(Score – E)	Ranks of (Score – E)
Students scoring 1	8	−2	22.5
Students scoring 2	14	−1	9.5
Students scoring 3	13	0	0
Students scoring 4	4	+1	9.5
Students scoring 5	0	+2	22.5

Sum of negative ranks: $T- = (14 \times 9.5) + (8 \times 22.5) = 133 + 180 = 313$
Sum of positive ranks: $T+ = (4 \times 9.5) = 38$
The smaller of the two ranks is the test statistic: $T = 38$.

The ties (where the scores equal the expected value and so Score $- E = 0$) contribute nothing to the comparison and are ignored. The ranks of the differences are given from 1 = the lowest to 26 = the highest. When there are ties between the ranks of the differences, each one is given the mean of the ranks. Hence in this case there are 18 students with a difference from expected of 1 (14 with -1 and 4 with $+1$). The mean of ranks 1 to 18 is 9.5, so each of these students is given 9.5. There are also eight students with a difference from the expected value of 2. The mean of ranks 19 to 26 is 22.5, so each of these students is given a rank of 22.5. Adding up the ranks gives a value for $T-$ of 313 and for $T+$ of 38. The smaller value is $T+$, so the T value to use is 38.

Using SPSS

Enter your data values into one column (here named, say, **scores**) and the expected mean values (here, 3) as a second column (named, say **expected**). Next, click on the **Analyze** menu, move onto the **Nonparametric Tests** bar, onto **Legacy Dialogs** and click on the **two Related Samples** bar. SPSS will produce the **Two Related Samples Tests** dialogue box. Put the columns to be compared into the **Test Pair(s) List** box, making sure the **Wilcoxon** test type is ticked. To get descriptive statistics and quartiles, go into **Options** and click on both **Descriptives** and **Quartiles**. The completed boxes and data are shown in the following screenshot.



	score	expected
1	1.00	3.00
2	1.00	3.00
3	1.00	3.00
4	1.00	3.00
5	1.00	3.00
6	1.00	3.00
7	1.00	3.00
8	1.00	3.00
9	2.00	3.00
10	2.00	3.00
11	2.00	3.00
12	2.00	3.00
13	2.00	3.00
14	2.00	3.00
15	2.00	3.00
16	2.00	3.00
17	2.00	3.00
18	2.00	3.00
19	2.00	3.00
20	2.00	3.00
21	2.00	3.00
22	2.00	3.00
23	3.00	3.00
24	3.00	3.00
25	3.00	3.00
26	3.00	3.00
27	3.00	3.00
28	3.00	3.00
29	3.00	3.00
30	3.00	3.00

Test Pairs:

Pair	Variable1	Variable2
1	[score]	[expecte...
2		

Test Type
☑ Wilcoxon
☐ Sign
☐ McNemar
☐ Marginal Homogeneity

Statistics
☑ Descriptive ☑ Quartiles

Missing Values
● Exclude cases test-by-test
○ Exclude cases listwise

Finally, click on **Continue** to get back to the main dialogue box and onto **OK** to run the test. SPSS will come up with the following results.

➡