

Chapter 13

The *t*-test: comparing two groups of unrelated/uncorrelated scores

- The uncorrelated or unrelated *t*-test is used to calculate whether the means of two sets of scores are significantly different from each other. It is the most commonly used version of the *t*-test.
- The unrelated *t*-test is used when the two sets of scores come from two different samples of people.
- (Turn to the previous chapter on the related *t*-test if your scores come from just one set of people or if you have employed a matching procedure.)
- Data entry for related and unrelated variables is very different in SPSS so take care to plan your analysis before entering your data in order to avoid problems and unnecessary work.
- SPSS procedures for the unrelated *t*-test are very useful since they include an option for its calculation when the variances of the two samples of scores are significantly different from each other. This is very rarely mentioned in statistics textbooks but is valuable nonetheless. If you have more than two sets of scores to compare, then turn to Chapter 20 on the unrelated analysis of variance.

We will illustrate the computation of an unrelated *t*-test with the data in Table 13.1 which shows the emotionality scores of 12 children from two-parent families and 10 children from single-parent families (ISP Table 13.8). In SPSS this sort of *t*-test is called an independent samples *t*-test.

Table 13.1 Emotionality scores in two-parent and lone-parent families

Two-parent family X_1	Lone-parent family X_2
12	6
18	9
14	4
10	13
19	14
8	9
15	8
11	12
10	11
13	9
15	
16	

13.1 Unrelated *t*-test

Quick summary

Enter data

Var00001 contains code for the two groups

Var00002 consists of the dependent variable

Statistics

Compare Means

Independent-Samples *T* Test . . .

Grouping variable (e.g. var00001)



Define Groups . . .

1

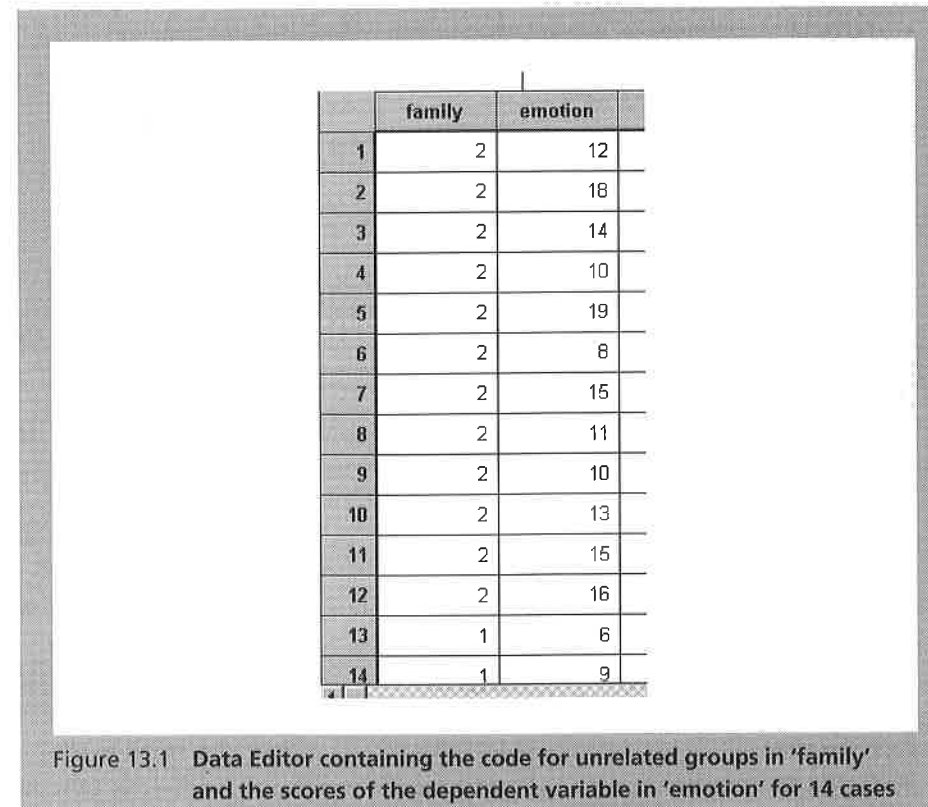
2

Continue

Dependent or test variable (e.g. var00002)



OK



	family	emotion
1	2	12
2	2	18
3	2	14
4	2	10
5	2	19
6	2	8
7	2	15
8	2	11
9	2	10
10	2	13
11	2	15
12	2	16
13	1	6
14	1	9

Figure 13.1 Data Editor containing the code for unrelated groups in 'family' and the scores of the dependent variable in 'emotion' for 14 cases

- Take a good look at Figure 13.1. Notice that there are two columns of data. The second column ('emotion') consists of the 22 emotionality scores from BOTH groups of children. The data are not kept separate for the two groups. In order to identify to which group the child belongs, the first column ('family') contains lots of 1s and 2s. These indicate, in our example, children from a lone-parent family (they are the rows with 1s in 'family') and children from two-parent families (they are the rows with 2s in 'family'). Thus a single column is used for the dependent variable (in this case, emotionality, 'emotion') and another column for the independent variable (in this case, type of family, 'family'). So each row is a particular child, and their independent variable and dependent variable scores are entered in two separate columns in the Data Editor.
- Save this data set, as it is used again in Chapter 18.
- Select 'Statistics' on the menu bar near the top of the window, which produces a drop-down menu (Figure 1.9).

- Select 'Compare Means' from this drop-down menu, which opens a second drop-down menu.
- Select 'Independent-Samples T Test ...', which opens the 'Independent-Samples T Test' dialog box (Figure 13.2).
- Select 'family' and the ► button beside 'Grouping Variable:', which puts 'family' in this box.
- Select 'Define Groups ...', which opens the 'Define Groups' sub-dialog box (Figure 13.3).
- Type '1' in the box beside 'Group 1:'.
- Select box beside 'Group 2:' and type '2'.
- Select 'Continue', which closes the 'Define Groups' sub-dialog box.

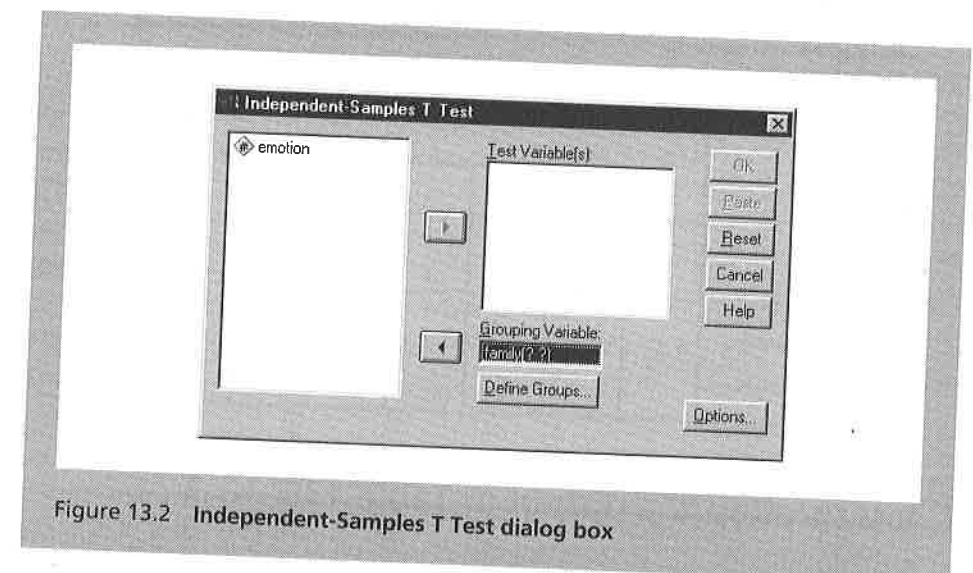


Figure 13.2 Independent-Samples T Test dialog box

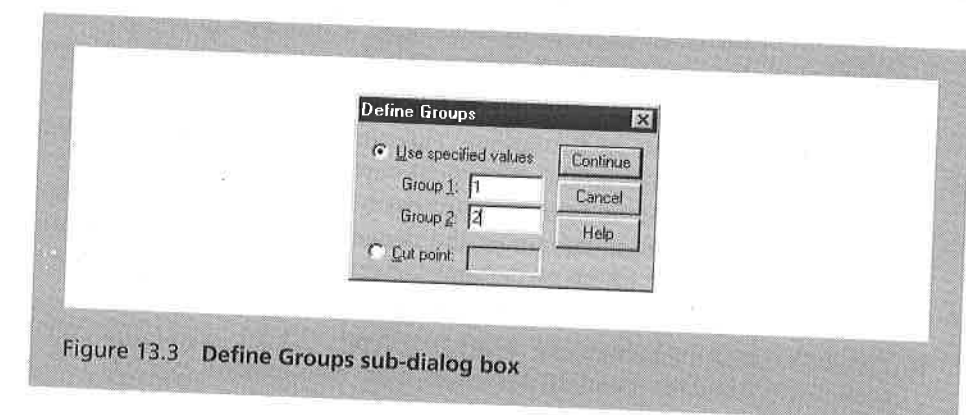


Figure 13.3 Define Groups sub-dialog box

- Select 'emotion' and the ► button beside 'Test Variable[s]:', which puts 'emotion' in this box.
- Select 'OK', which closes the 'Independent-Samples T Test' dialog box and the Data Editor window and which displays the output shown in Table 13.2.

Table 13.2 Unrelated *t*-test output

Group Statistics					
	FAMILY	N	Mean	Std. Deviation	Std. Error Mean
EMOTION	1	10	9.50	3.10	.98
	2	12	13.42	3.37	.97

Paired Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
			t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	F	Sig.						Lower	Upper
EMOTION	.212	.650	-2.813	20	.011	-3.92	1.39	-6.82	-1.01
			-2.836	19.768	.010	-3.92	1.38	-6.80	-1.03

13.2 Interpreting the output in Table 13.2

The output for the uncorrelated/unrelated *t*-test on SPSS is particularly confusing even to people with a good knowledge of statistics. The reason is that there are two versions of the uncorrelated/unrelated *t*-test. Which one to use depends on whether or not there is a significant difference between the (estimated) variances for the two groups of scores.

- Examine the first table of the output. This contains the means and standard deviations of the scores on the dependent variable (emotionality) of the two groups. Notice that an additional figure has been added by the computer to the name of the column containing the dependent variable. This additional figure indicates which of the two groups the row refers to. If you had labelled your values, these value labels would be given in the table.
- For children from two-parent families ('family 2') the mean emotionality score is 13.42 and the standard deviation of the emotionality scores is 3.37. For

the children of lone-parent families ('family 1') the mean emotionality score is 9.50 and the standard deviation of emotionality is 3.10 respectively.

- In the second table, read the line 'Levene's Test for Equality of Variances'. If the probability value is statistically significant then your variances are UNEQUAL. Otherwise they are regarded as equal.
- Levene's test for equality of variances in this case tells us that the variances are equal because the *p* value of .650 is not statistically significant.
- Consequently, you need the row for 'Equal variances assumed'. The *t* value, its degrees of freedom and its probability are displayed. The *t* value for equal variances is -2.813, which with 20 degrees of freedom has an exact two-tailed significance level of .011.
- Had Levene's test for equality of variances been statistically significant (i.e. .05 or less), then you should have used the second row of the output which gives the *t*-test values for unequal variances.

13.3 Reporting the results of Table 13.2

- We could report the results of this analysis as follows: "The mean emotionality scores of children from two-parent families ($M = 13.42$, $SD = 3.37$) is significantly higher ($t = -2.81$, $df = 20$, two-tailed $p = .011$) than that of children in lone-parent families ($M = 9.50$, $SD = 3.10$)."
- It is unusual to see the *t*-test for unequal variances in psychological reports. Many psychologists are unaware of its existence. So what happens if you have to use one? In order to clarify things, we would write: "Because the variances for the two groups were significantly unequal ($F = 8.43$, $p < .05$), a *t*-test for unequal variances was used . . .".
- If you prefer to use the confidence intervals in Table 13.2, you might write: "The difference between the emotionality scores for the children from two-parent families ($M = 13.42$, $SD = 3.37$) and lone-parent families ($M = 9.50$, $SD = 3.10$) is -3.92. The 95% confidence interval for this difference is -6.82 to -1.01. Since this interval does not include 0.00, the difference is statistically significant at the two-tailed 5% level."

of scores on a variable. It takes a positive value if the values are skewed to the left and a negative value if they are skewed to the right.

Kurtosis – an index of how much steeper or flatter the distribution of scores on the variable is compared to the normal distribution. It takes a '+' sign for steep frequency curves and a '-' sign for flat curves.

Standard deviation (estimate) – this is a measure of the amount by which scores differ on average from the mean of the scores on a particular variable. Its method of calculation involves unusual ways of calculating the mean. In SPSS the standard deviation is calculated as an estimate of the population standard deviation. It is an index of the variability of scores around the mean of a variable. Some authors call this the sample standard deviation.

Variance (estimate) – this is a measure of the amount by which scores on average vary around the mean of the scores on that variable. It is the square of the standard deviation and is obviously therefore closely related to it. It is also always an estimate of the population variance in SPSS. Some authors call this the sample variance. Like standard deviation, it is an index of the variability of scores around the mean of a variable but also has other uses in statistics. In particular, it is the standard unit of measurement in statistics.

Range – the numerical difference between the largest and the smallest scores obtained on a variable. It is a single number.

Minimum (score) – the value of the lowest score in the data for a particular variable.

Maximum (score) – the value of the highest score in the data for a particular variable.

Standard error (SE mean) – the average amount by which the means of samples drawn from a population differ from the population mean. It is calculated in an unusual way. Standard error can be used much like standard deviation and variance as an index of how much variability there is in the scores on a variable.

Chapter 4

Shapes of distributions of scores

- *It is important to study the shape of the distribution of scores on a variable. Ideally for most statistical techniques, a distribution should be symmetrical and normally distributed (bell-shaped).*
- *Researchers should be wary of very asymmetrical (skewed) distributions and distributions which contain a few unusually high or low scores (outliers). Histograms, for example, can be used to detect asymmetry and outliers.*

We will compute a frequency table and histogram of the extraversion scores of the 50 airline pilots shown in Table 4.1 (ISP Table 4.1).

Table 4.1 Extraversion scores of 50 airline pilots

3	5	5	4	4	5	5	3	5	2
1	2	5	3	2	1	2	3	3	3
4	2	5	5	4	2	4	5	1	5
5	3	3	4	1	4	2	5	1	2
3	2	5	4	2	1	2	3	4	1

4.1 Frequency tables

Quick summary

Enter the data

Statistics

Summarize

FrequenciesSelect variable ► buttonOK

- As described in Chapter 1, enter the extraversion scores in the first column of the Data Editor window. Label the variable (e.g. 'extrav') and remove the two zero decimal places.
- Save the data, as described in Chapter 1, in a system file called 'ext.sav'.
- Select 'Statistics', which produces a drop-down menu (Figure 1.9).
- Select 'Summarize', which displays a second drop-down menu (Figure 1.9).
- Select 'Frequencies . . .', which opens the Frequencies dialog box (Figure 2.7).
- Select 'extrav' (if you have only one variable this will already be highlighted for you) and then the ► button, which puts 'extrav' in the 'Variable[s]:' text box.
- Select 'OK', which closes the Frequencies dialog box and the Data Editor window and which displays the output shown in Table 4.2 in the Viewer window.

Table 4.2 Frequency table of extraversion scores of 50 airplane pilots as produced by Frequencies

Statistics**EXTRAV**

N	Valid	50
	Missing	0

EXTRAV

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	7	14.0	14.0	14.0
	2	11	22.0	22.0	36.0
	3	10	20.0	20.0	56.0
	4	9	18.0	18.0	74.0
	5	13	26.0	26.0	100.0
	Total	50	100.0	100.0	

4.2 Interpreting the output in Table 4.2

- Column 1: The five categories of extraversion scores are listed in the left-most column of numbers. They are 1, 2, 3, 4 and 5.
- Column 2: The frequency of each of these categories is to be found here. Thus there are 10 cases with an extraversion value of 3.
- Column 3: The percentage of cases in each category for the sample as a whole. This includes any values that you may have defined as missing values. Thus 20% of the cases in the sample are in the 3 category.
- Column 4: The percentage of cases in each category for the sample excluding any missing values. Since there were no values defined as missing in our example, columns 3 and 4 are identical.
- Column 5: The cumulative percentage excluding any missing values is found in the final column of numbers. Thus 74% of the cases had a value of 4 or less.

4.3 Reporting the output of Table 4.2

Table 4.3 One style of reporting output in Table 4.2

Extraversion score	Frequency	Percentage frequency	Cumulative percentage frequency
1	7	14.0	14.0
2	11	22.0	36.0
3	10	20.0	56.0
4	9	18.0	74.0
5	13	26.0	100.0

Notice that we omitted some of the confusion of detail in Table 4.2. Tables and diagrams need to clarify the results.

4.4 Histograms

Quick summary

Enter the data or retrieve file

Graphs

Histogram

Select variable (e.g. 'extrav') ► button

OK

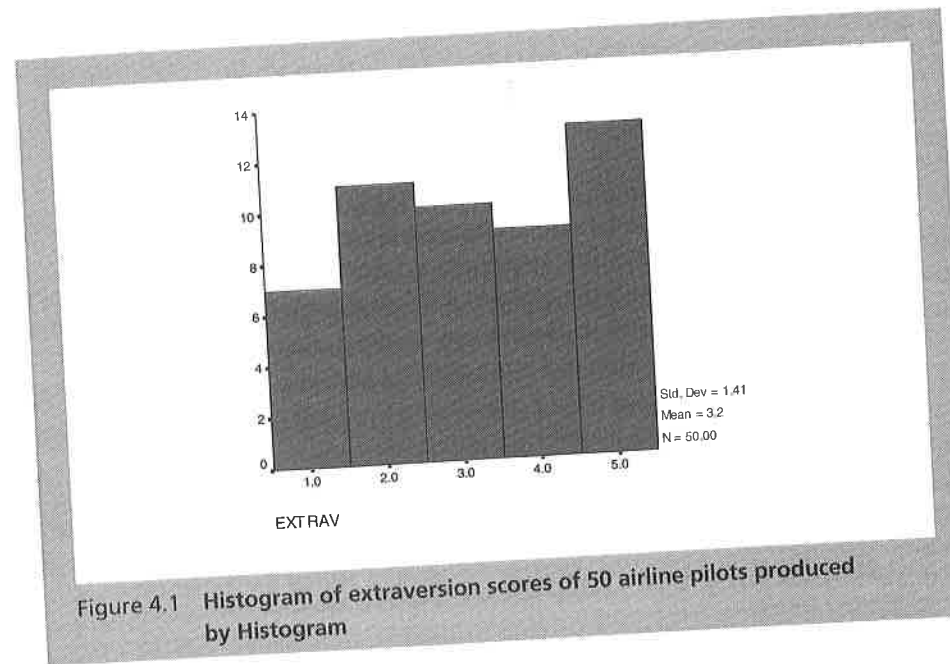


Figure 4.1 Histogram of extraversion scores of 50 airline pilots produced by Histogram

- Select 'Graphs', which produces a drop-down menu (Figure 2.8).
- Select 'Histogram . . .', which displays the Histogram dialog box (Figure 2.15).
- Select 'extrav' (which will already be highlighted for you if you only have one variable) and then the ► button which puts 'extrav' in the 'Variable:' text box.
- Select 'OK', which closes the Histogram dialog box and the Data Editor window and which displays the output shown in Figure 4.1 in the Viewer window.
- This histogram can be edited by putting it in the Chart Editor (see section 2.14). Examples of the results of such editing can be seen in Figure 4.2.

4.5 Interpreting the output in Figure 4.1

- The horizontal axis of this histogram gives the values 1.0, 2.0, 3.0, 4.0 and 5.0. These are the five different values of extraversion scores in Table 4.1.
- The vertical axis gives frequencies of individuals or cases from 0 to 14 in steps of two.
- The heights of the bars tell you how frequent each extraversion score is in the data.
- Std.Dev = 1.41 is the value of the standard deviation. This is discussed again in Chapter 5 of this book.
- The average of the fifty extraversion scores is given by Mean = 3.2.
- N = 50.00 means that the number of scores = 50.

4.6 Reporting the output in Figure 4.1

- Although Figure 4.1 may be easily interpreted by you, it is not suitable for inclusion unmodified in your reports.
- The vertical axis needs to be labelled 'f' or 'frequencies'.
- The horizontal axis needs to be labelled 'extraversion scores'.
- Figure 4.2 contains the additional information needed.

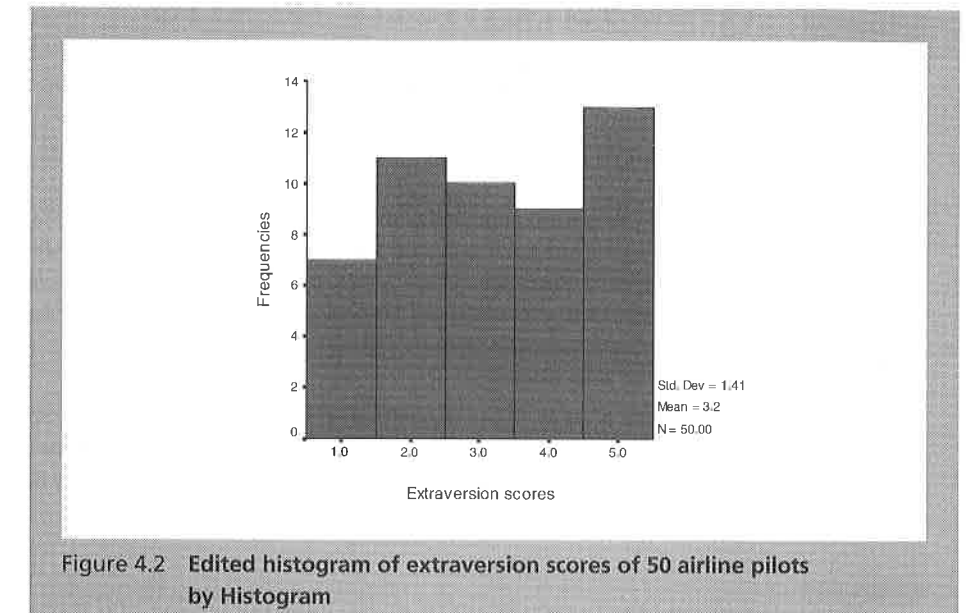


Figure 4.2 Edited histogram of extraversion scores of 50 airline pilots by Histogram

Chapter 5

Standard deviation: the standard unit of measurement in the social sciences

- Basically, standard deviation is an index of how much scores deviate (differ) 'on average' from the average of the set of scores of which they are members. In other words, standard deviation is an index of the amount of variability of scores around the mean of the scores. Standard deviation can also be used in order to turn scores on very different variables into *z*- or 'standard scores' which are easily compared and summed.
- Standard deviation always takes a positive value BUT researchers write of a number of standard deviations above the mean (i.e. '+' relative to the mean) or a number of standard deviations below the mean (i.e. '-' relative to the mean).

We shall illustrate the computation of the standard deviation and *z*-scores with the nine age scores shown in Table 5.1 (based on *ISP* Table 5.1).

Table 5.1 Data for the calculation of standard deviation

Age	20	25	19	35	19	17	15	30	27
-----	----	----	----	----	----	----	----	----	----

5.1 Standard deviation

Quick summary

Statistics

Summarize

Descriptives

Select variables ►

Options

Std deviations

Continue

OK

- Enter the data into the Data Editor, label the variable (e.g. 'age') and remove the two zero decimal places.
- Select 'Statistics', which produces a drop-down menu (Figure 1.9).
- Select 'Summarize', which displays a second drop-down menu (also Figure 1.9).

either

- Select 'Frequencies ...' and follow the procedure described in Chapter 3, selecting 'Std deviation' in the Frequencies: Statistics sub-dialog box (Figure 3.1).

or

- Select 'Descriptives ...', which opens the Descriptives dialog box (Figure 1.10).

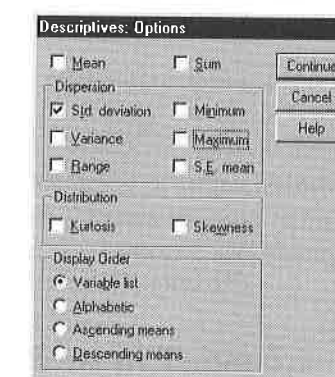


Figure 5.1 Descriptives: Options sub-dialog box

- Select 'age' and then the ► button, which puts 'age' in the 'Variable[s]:' text box.
- Select 'Options . . .', which opens the Descriptives: Options sub-dialog box (Figure 5.1).
- Select 'Std deviations' and de-select 'Mean', 'Minimum' and 'Maximum' by clicking off these options (if you only want standard deviations). Normally you will find the mean and minimum plus maximum useful.
- Select 'Continue', which closes the Descriptives: Define sub-dialog box.
- Select 'OK', which closes the Descriptives dialog box and Data Editor window and which displays the output shown in Table 5.2 in the Viewer window.

Table 5.2 Standard deviation of nine ages as produced by Descriptives

Descriptive Statistics		
	N	Std. Deviation
AGE	9	6.6521
Valid N (listwise)	9	

5.2 Z-scores

Quick summary

Statistics

Summarize

Descriptives

Select 'Save standardized values as variables'

Select variables ►

Options

Std deviations

Continue

OK

	age	z:age	Var
1	20	-.45099	
2	25	.30066	
3	19	-.60132	
4	35	1.80395	
5	19	-.60132	
6	17	-.90198	
7	15	-1.20263	
8	30	1.05230	
9	27	.60132	

Figure 5.2 z-scores produced by Descriptives and displayed in the Data Editor

This involves just one extra step:

- To obtain z-scores we simply select 'Save standardized values as variables' in the Descriptives dialog box (Figure 1.10), which will put the z-scores in the next column in the Data Editor entitled 'z:age' as shown in Figure 5.2.

In other words, for each of your scores the z-score has been calculated and placed in a separate column in your data window. This new 'variable' will be lost unless you save your data before ending the session.

5.3 Interpreting the output in Table 5.2

- The standard deviation of the nine ages is 6.65. This is the 'average' amount by which each score deviates from the mean score in the set. Standard deviation is quite a complex concept so refer to ISP Chapter 5 if you have any difficulties or another textbook.
- The standard deviation is based on nine valid cases. In other words, the computer has not found any values of age which you previously have defined as 'missing' or to be ignored (Chapter 15).

5.4 Reporting the output in Table 5.2

- The standard deviation of just one variable can easily be mentioned in the text of your report: "It was found that the standard deviation of age was 6.65 years (N = 9)."

- However, it is more likely that you would wish to record the standard deviation alongside other statistics such as the mean and range, as illustrated in Table 5.3. You probably wish to include these statistics for other numerical score variables that you have data on.

Table 5.3 The sample size, mean, range and standard deviations of age, IQ and verbal fluency

	N	Mean	Range	Standard deviation
Age	9	23.00	20.00	6.65
IQ	9	122.17	17.42	14.38
Verbal fluency	9	18.23	4.91	2.36

5.5 Other features

'Descriptives' contains a number of statistical calculations which are easily selected. These different statistical concepts are briefly explained at the end of Chapter 3.

Mean
Sum
Standard deviation (estimate)
Range
Minimum (score)
Maximum (score)
Standard error (S.E. mean)
Kurtosis
Skewness

Chapter 6

Relationships between two or more variables: diagrams and tables

- A great deal of research explores the relationship between two or more variables. While the univariate (single-variable) statistical procedures described so far have their place in the analysis of practically any data, many research questions require methods which help assess the interrelationships or correlations between different variables in addition.
- As with univariate statistics, a thorough statistical analysis of data requires an exploration of the basic trends in the data by the use of tables and diagrams. These methods include cross-tabulation tables, compound histograms (clustered bar charts) and scattergrams. Most of these are familiar to many of us.
- Care has to be taken to make sure that the tables and diagrams you obtain are useful and communicate well. In particular, ensure that your data for cross-tabulation tables and compound histograms (clustered bar charts) only contain a small number of different data values. If they do not, SPSS will produce massive tables and dense unreadable graphs and diagrams.
- Scattergrams, on the other hand, also work well when you have many different values for the scores on your variables.
- Remember that the effective presentation of basic descriptive statistics requires that the researcher considers their tables and diagrams carefully. Often the researcher will have to 'tweak' things to make the finished tables and graphs effective forms of communication. That this can be done quickly is one of the advantages of using SPSS.

We will illustrate the drawing up of a cross-tabulation table and compound bar chart with the data shown in Table 6.1 (ISP Table 6.4). This shows the number of men and women in a study who have or have not been previously hospitalised.