4: Parametric tests of Differences Between Means

4.1. The Independent *t*-test

The independent t-test is used in experiments in which there are two conditions and different subjects have been used in each condition. For example, a psychologist might be interested in whether phobic responses are specific to a particular object, or whether they generalise to other, perceptually similar, objects. Twenty-four spider-phobes were used in all: 12 were exposed to a real tarantula spider and their subsequent anxiety was measured and the remaining 12 were shown only a toupee (a wig) that was perceptually similar to the spider (i.e. hairy and round). Likewise, their anxiety was measured. When we have collected data using different subjects in each group, we need to input the data using a coding variable (see Handout 1). So, the spreadsheet will have two columns of data. The first column is a coding variable (called something like group), which, in the case of the *t*-test, will have two codes (for convenience I suggest 0 = Toupee group, and 1 = real spider group). The second will have values for the dependent variable (anxiety). The data are in Table 4.1 in which the group codes are shown (rather than the group names). When you enter the data into SPSS remember to tell the computer that a code of 0 represents the group that were shown the toupee, and that a code of 1 represents the group that saw the real spider.

Table 4.1: Data for spider experiment

Subject	Group	Anxiety
1	0	30
2	0	35
3	0	45
4	0	40
5	0	50
6	0	35
7	0	55
8	0	25
9	0	30
10	0	45
11	0	40
12	0	50
13	1	40
14	1	35
15	1	50
16	1	55
17	1	65
18	1	55
19	1	50
20	1	35
21	1	30
22	1	50
23	1	60
[illegible]	[illegible]	[illegible]

4.1.1. Running the Analysis

First, we would run some exploratory analysis on the data (see Handout 2). Next we need to access the main dialogue box by using the **Analyze Þ Compare Means Þ Independent-Samples T Test …** menu pathway (see Figure 4.1). Once the dialogue box is activated, select the dependent variable from the list (click on **anxiety**) and transfer it to the box labelled *Test Variables* by clicking on . If you want to carry out t-tests on several dependent variables then you can select other dependent variables and transfer them to the variable list. However, there are good reasons why it is not a good idea to carry out lots of tests (you'll have to wait until your second year to find out why).

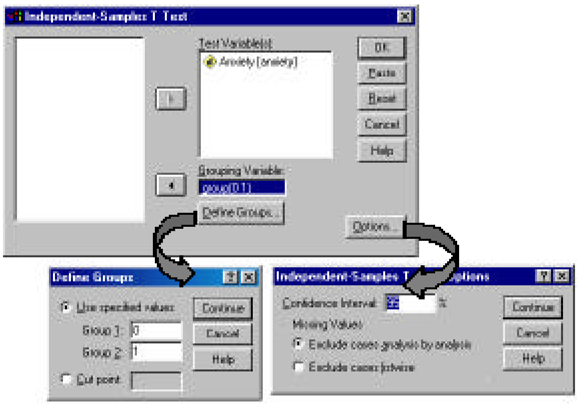


Figure 4.1: Dialogue boxes for the independent means *t*-test

Next, we need to select an independent variable (the grouping variable). In this case, we need to select **group** and then transfer it to the box labelled *Grouping variable*. When your grouping variable has been selected the Define Groups... button will become active and you should click on it to activate the *define groups* dialogue box. SPSS need to know what numeric codes you assigned to your two groups, and there is a space for you to type the codes. In this example, we coded our toupee group as 0, and our real group as 1 and so these are the codes that we type. Alternatively you can specify a *cut point* in which case SPSS will assign all cases greater than or equal to that value to one group and all the values below the cut point to the second group. This facility is useful if you are testing different groups of subjects based on something like a median split. So, you might want to classify people as spider-phobes or non-spider-phobes, and so you measure their score on a spider phobia questionnaire and calculate the median. You then classify anyone with a score above the median as a phobic, and those below the median as non-phobes. Rather than recoding all of your subjects and creating a coding variable, you would simply type the median value in the box labelled *cut point*. When you have defined the groups, click on Continue to return to the main dialogue box. If you click on Options... then another dialogue box appears that gives you the same options as for the dependent t-test. To run the analysis click on OK.

4.1.2. Output from the Independent t-test

The output from the independent t-test contains only two tables. The first table provides summary statistics for the two experimental conditions. From this table, we can see that both groups had 12 subjects in (column labelled N), the group who saw the toupee had a mean anxiety of 40, with a standard deviation of 9.29. What's more, the standard error of that group (the standard deviation of the sampling distribution) is 2.68 ($SE = \frac{s}{\sqrt{N}} = \frac{9.29}{\sqrt{12}} = 2.68$). In addition, the table tells us that the average anxiety level in participants who were shown a real spider was 47, with a standard deviation of 11.03, and a standard error of 3.18 ($SE = \frac{s}{\sqrt{N}} = \frac{11.03}{\sqrt{12}} = 3.18$).

Group Statistics

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Anxiety	Picture	12	40.0000	9.2932	2.6827
	Real Spider	12	47.0000	11.0289	3.1838

4.1.2.1. The assumption of Homogeneity of Variance

The second table of output contains the main test statistics. The first thing to notice is that there are two rows containing values for the test statistics, one row is labelled *equal variances assumed*, while the other is labelled *equal variances not assumed*. The t-test assumes that the variances in experimental groups are roughly equal. Well, in reality there are adjustments that can be made in situations in which the variances are unequal, and the rows of the table relate to whether or not this assumption has been broken. How do we know whether this assumption has been broken?

Well, we could just look at the values of the variances and see whether they are similar (for example, we know that the standard deviations of the two groups are 9.29 and 11.03, if we square these values then we get the variances). However, this would be a very subjective measure and probably prone to academics thinking 'ooh look, the variance of group 1 is only 3000 times larger than the variance of group 2: that's roughly equal'. Fortunately, there is a test that can be performed to see whether the variances are different enough for us to be concerned. Levene's test (as it is known) can be conceptualised as similar to a t-test in that it tests the hypothesis that the variances in the two groups are equal (i.e. the difference between the variances is zero). Therefore, if Levene's test is significant at $p \leq 0.05$ then we can conclude that the null hypothesis is incorrect and that the variances are significantly different (and so the assumption of homogeneity of variances has been violated). If, however, Levene's test is nonsignificant (i.e. $p > 0.05$) then we must accept the null hypothesis that the difference between the variances is zero and so the variances must be roughly equal (and so the assumption is tenable). For these data, Levene's test is nonsignificant (because $p = 0.386$, which is greater than 0.05) and so we should read the test statistics in the row labelled *equal variances assumed*. Had Levene's test been significant, then we would have read the test statistics from the row *labelled equal variances not assumed*.

4.1.2.2. The Main Analysis

Having established that the assumption of homogeneity of variances is met, we can move on to look at the t-test itself. We are told the mean difference ($\overline{X}_1 - \overline{X}_2 = 40 - 47 = -7$) and the standard error of the sampling distribution of differences, which is calculated using the lower half of the equation for the t-test ($\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{(9.29)^2}{12} + \frac{(11.03)^2}{12}} = \sqrt{7.19 + 10.14} = \sqrt{17.33} = 4.16$). The *t*-statistic itself is calculated by dividing the mean difference by the standard error of the sampling distribution of differences ($t = \frac{-7}{4.16} = -1.68$). The value of *t* is then assessed against the value of *t* you might expect to get when you have certain degrees of freedom. The degrees of freedom reflect the sample sizes and the number of samples taken. For the *t*-test, degrees of freedom are calculated by adding the two sample sizes and then subtracting the number of samples ($df = N_1 + N_2 - 2 = 12 + 12 - 2 = 22$). SPSS produces the exact significance value of *t*, and we are interested in whether this value is less than or greater than 0.05. In this case, the two-tailed value of *p* is 0.107, which is greater than 0.05 and so we would have to conclude that there was no significant difference between the means of these two samples. In terms of the experiment, we can infer that spider-phobes are made equally anxious by toupees as they are by real spiders.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Mean	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Anxiety	Equal variances assumed	.782	.386	-1.681	22	.107	-7.0000	4.1633	-15.6342	1.6342
	Equal variances not assumed			-1.681	21.385	.107	-7.0000	4.1633	-15.6486	1.6486

4.1.2.3. One-Tailed tests

Now, we use the two-tailed probability when we have made no specific prediction about the direction of our effect. For example, if we were unsure whether a real spider would induce more or less anxiety, then we would have to use a two-tailed test. However, often in research we can make specific predictions about which groups will have the biggest means. In this example, it is likely that we would have predicted that a real spider would induce greater anxiety than a toupee and so we predict that the mean of the real group would be greater than the mean of the toupee group. In this case, we can use a one-tailed test. Some students get very upset by the fact that SPSS produces only the two-tailed significance and are confused by why there isn't an option that can be selected to produce the one-tailed significance. The answer is simple: there is no need for an option because the one-tailed probability can be ascertained by dividing the two-tailed significance value by two. In this case, the two-tailed probability was 0.107, therefore the one-tailed probability is 0.054 ($\frac{0.107}{2}$). The one-tailed probability is still greater than 0.05 (albeit by a small margin) and so we would still have to conclude that spider-phobes were equally as anxious when presented with a real spider as spider-phobes who were presented with a toupee.

4.2. Graphing Repeated Measures

In weeks 1 and 2, we have seen how to enter data using a coding variable. This type of variable is used to define different groups of people (or different experimental conditions in which different subjects are used). However, in some experiments, we use the same group of subjects in several experimental conditions. This type of experimental design is called a repeated measures (or within-subjects) design. One of the advantages of this methodology is that individual differences are controlled across conditions. In the last example, the lack of significance could be due to large differences in the degree to which people in different conditions were phobic (for example, the toupee group may have contained highly phobic subjects and the spider condition relatively low level phobics). Imagine the psychologist repeated the experiment, but this time used the same subjects in the two conditions. Half of the subjects were exposed to the toupee first, and half to the spider (to eliminate order effects). The data are in Table 4.2 (and in fact the scores are identical to the previous example). The data are entered into the data editor differently now because the same subjects were used. In SPSS, each row of the spreadsheet represents a single subject and so with repeated measure designs, the data are arranged in columns (one representing the **toupee** condition and one representing the **real** condition).

Table 4.2: Data for spider example

Subject	Toupee (Anxiety score)	Real (Anxiety Score)
1	30	40
2	35	35
3	45	50
4	40	55
5	50	65
6	35	55
7	55	50
8	25	35
9	30	30
10	45	50
11	40	60
12	50	39

To plot a bar graph of these data simply select the error bar dialogue box by using the mouse to select **Graphs** ⇒ **Bar** **…** and then in the dialogue box (see) click on [Summaries of separate variables] and then [Define]. This process will bring up the main dialogue box (see Figure 4.2). Once the dialogue box is activated you simply select the two variables of interest (toupee and real) and then click on [OK].

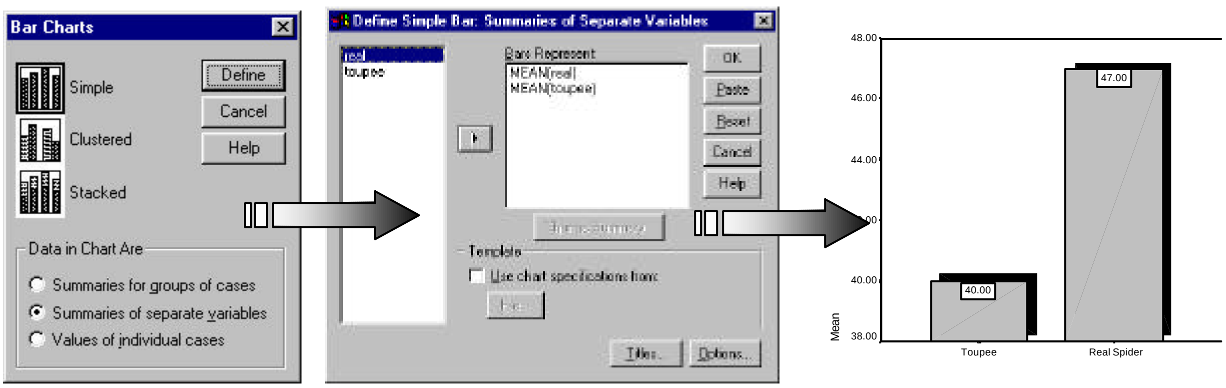


Figure 4.2: Graphing repeated measures variables.

4.3. The Compute Function

The *Compute* function allows us to carry out various functions on columns of data. Some typical functions might be adding the scores across several columns, taking the square root of all of the scores in a column, or taking the mean of several variables. To access the compute dialogue box, use the mouse to specify **Transform Þ Compute …** The resulting dialogue box is shown in Figure 4.3. In this dialogue box there is a list of functions on the right hand side, and a calculator-like keyboard in the centre. Most of the functions on the calculator are obvious but the most common are listed below. The results of any compute function will be produced in a new column in the spreadsheet and so the first thing to do is to type in a label for this new variable (in the box labelled *Target Variable*). If you type in a variable name that already exists in the spreadsheet then SPSS will tell you and ask you whether you want to replace this existing variable. If you respond with *Yes* then SPSS will replace the data in the existing column with the result of the compute function. If you respond with *No* then nothing will happen and you will need to rename the *Target variable*. The box labelled *Numeric Expression* is the space where arithmetic commands can be typed (I've called this space the command area). You can enter variable names into the command area by selecting the variable required from the variable list and then clicking on . Likewise, you can select certain functions from the list of available functions and enter them into the command area by clicking on .

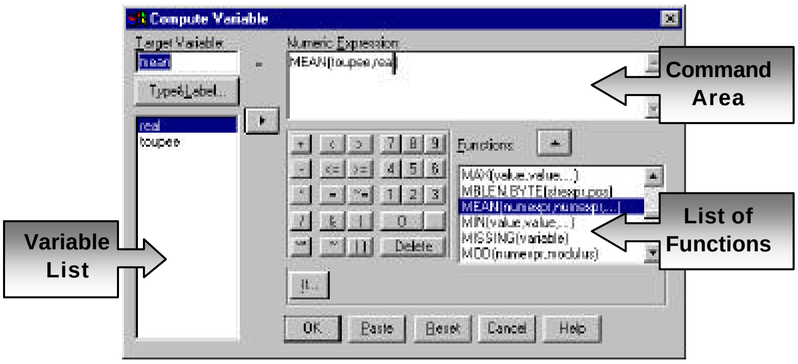


Figure 4.3: Dialogue box for the Compute function

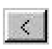
Addition: this button places a plus sign in the command area. For example, 'picture + real' creates a column where each row contains the score from the column labelled *picture* added to the score from the column labelled *real* (e.g. for subject 1: 30 + 40 = 70)

Subtraction: this button places a minus sign in the command area. For example, 'picture – real' creates a column where each row contains the score from the column labelled *real* subtracted from the score in the column labelled *picture* (e.g. for subject 1: 30 – 40 = -10).

Multiply: this button places a multiplication sign in the command area. For example, 'picture * real' creates a column that contains the score from the column labelled *picture* multiplied by the score in the column labelled *real* (e.g. for subject 1: 30 × 40 = 1200).

Divide: this button places a division sign in the command area. For example, 'picture/real' creates a column that contains the score from the column labelled *picture* divided by the score from the score in column labelled *real* (e.g. for subject 1: 30 ÷ 40 = 0.75).

Exponentiation: this button is used to raise the preceding term by the power of the proceeding term. So, 'picture **2' creates a column that contains the scores in the *picture* column raised to the power of 2 (i.e. the square of each number in the picture column — for subject 1 = $(30)^2$ = 900) Likewise, 'picture**3' creates a column with values of *picture* cubed.



Less than: This operation is usually used for 'include case' functions. If you click on  button, you will see a dialogue box that allows you to select certain cases to carry out the operation on. So, if you typed 'picture < 30', then SPSS would carry out the compute function only for those subjects whose anxiety in the picture condition was less than 30 (i.e. if anxiety was 29 or less).

<= **Less than or equal to**: This operation is the same as above except that when used it will include cases that are exactly 30 as well.

> **More than**: This operation is the opposite of less than and is generally used include cases above a certain value. So, if you clicked on If... and then typed 'picture > 30' then SPSS will carry out any analysis only on cases for which anxiety in the picture condition was greater than 30 (i.e. 31 and above)

>= **More than or equal to**: This operation is the same as above but will include cases that are exactly 30 as well.

= **Equal to**: You can use this operation to included cases for which subjects have a specific value. For example, 'picture = 30' typed in the *if* dialogue box will ensure that only cases that have a value of 30 for the picture condition are included.

~= **Not equal to**: This operation will include all cases except those will a specific value. So, 'picture ~= 30' will include all cases except those that had an anxiety score of 30 in the picture condition.

Type&Label... **Variable Type**: This button opens a dialogue box that allows you to give the new variable a full name, and to specify what type of variable it is (e.g. numeric). Mostly you will not need to use this button.

Table 4.3: Some useful *Compute* functions

Function	Name	Example Input	Output
Mean(?,?, ..)	Mean	Mean(picture, real)	For each row, SPSS calculates the average of the variables picture and real
SD(?,?, ..)	Standard Deviation	SD(picture, real)	Across each row, SPSS calculates the standard deviation of the values in the columns labelled picture and real.
SUM(?,?, ..)	Sum	SUM(picture, real)	For each row, SPSS adds the values in the columns labelled picture and real.
SQRT(?)	Square Root	SQRT(picture)	Produces a column that contains the square root of each value in the column labelled picture. Useful for transforming skewed data or data with heterogeneous variances.
ABS(?)	Absolute Value	ABS(picture)	Produces a variable that contains the absolute value of the values in the column labelled picture (absolute values are ones where the ± signs are ignored: so –5 becomes +5 and +5 stays as +5).
LG10(?)	Base 10 Logarithm	LG10(picture)	Produces a variable that contains the logarithmic (to base 10) values of picture. This is useful for transforming positively skewed data.
Normal(stddev)	Normal Random Numbers	Normal(5)	Produces a variable containing pseudo-random numbers from a normal distribution with a mean of 0 and a standard deviation of 5)

Some of the most useful functions are listed in Table 4.3, which shows the standard form of the function, the name of the function, an example of how the function can be used and what SPSS would output if that example were used. There are several basic functions for calculating means, standard deviations and sums of columns. There are also functions such as the square root and logarithm functions that are useful for transforming data that are skewed (see Howell 1997, chapter 11 for a discussion of data transformation). For the interested reader, the SPSS base systems user's guide has details of all of the functions available through the *Compute* dialogue box (Norušis, 1997; see also SPSS Inc., 1997).

4.3.1. Exercise: Calculate the Mean for each Subject

Now we know something about the *Compute* function, we can use this function to produce an average anxiety score across the two conditions (this might be useful for the experimenter to gauge which subjects were most anxious overall). We can find out the average across conditions by using the *Mean(?,?)* function. Access the main *Compute* dialogue box by using the **Transform Þ Compute …** menu path. Enter the name **mean** into the box labelled *Target variable* and then scroll down the list of functions until you find the one called *Mean(numexpr, numexpr,..)*. Highlight this function and transfer it to the command area by clicking on . When the command is transferred, it will appear in the command area as Mean(?,?) and the question marks should be replaced with variable names (which can be typed manually or transferred from the variable list). So replace the first question mark with the variable **toupee**, and the second one with the variable **real**. The completed dialogue box should look like the one in Figure 4.3. If you click on OK a new column will appear in the data editor containing the average anxiety level for each subject.

4.4. The Dependent *t*-test

4.4.1. Running the Analysis

Using our spider data (Table 4.2), we have twelve spider phobics who were exposed to a toupee (**toupee**) and on a separate occasion a real life tarantula (**real**). Their anxiety was measured in each condition. I have already described how the data are arranged, and so we can move straight onto doing the test itself. First, we need to access the main dialogue box by using the **Analyze Þ Compare Means Þ Paired-Samples T Test …** menu pathway (Figure 4.4). Once the dialogue box is activated, select two variables from the list (click on the first variable with the mouse and then the second). The first variable you select will be named as *Variable 1* in the box labelled *Current Selections*, and the second variable you select appears as *Variable 2*. When you have selected two variables, transfer them to the box labelled *Paired Variables* by clicking on ▸. If you want to carry out several *t*-tests then you can select another pair of variables, transfer them to the variable list, and then select another pair and so on. In this case, we want only one test. If you click on Options... then another dialogue box appears that gives you the chance to change the width of the confidence interval that is calculated. The default setting is for a 95% confidence interval and this is fine, however, if you want to be stricter about your analysis you could choose a 99% confidence interval but this runs a higher risk of failing to detect a genuine effect (a Type II error). To run the analysis click on OK.

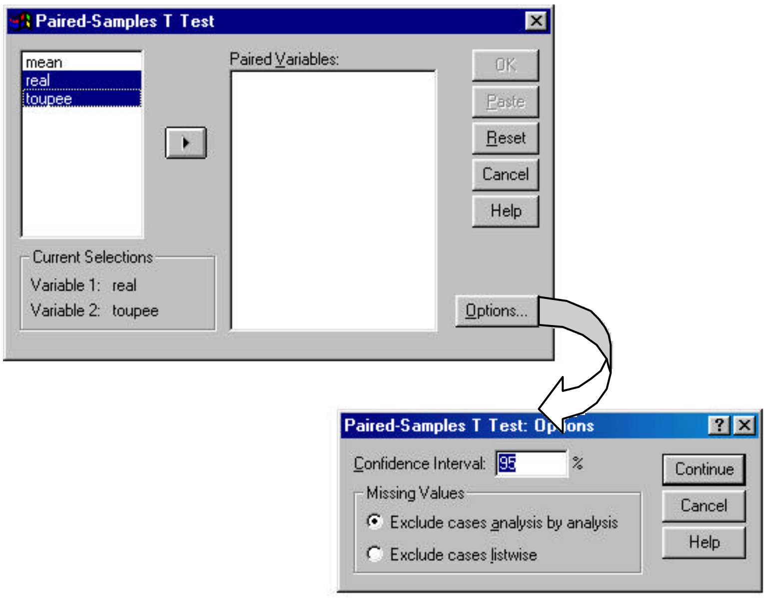


Figure 4.4: Main dialogue box for Paired-Samples *t*-test

4.4.2. Output from SPSS

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Real Spider	47.0000	12	11.0289	3.1838
	Toupee	40.0000	12	9.2932	2.6827

The resulting output produces three tables. In the first table we are told the mean, the number of subjects (*N*) and the standard deviation of each sample. In the final column we are told the standard error, which is the sample standard deviation divided by the square root of the sample size $\left(SEM = \frac{s}{\sqrt{N}}\right)$, so for the picture condition $SEM = \frac{9.2932}{\sqrt{12}} = \frac{9.2932}{3.4641} = 2.68$. SPSS also shows the Pearson correlation between the two conditions. When repeated measures are used it is possible that the experimental conditions will correlate (because the data in each condition comes from the same people and so there could be some constancy in their responses). SPSS provides the value of Pearson's *r* and the two-tailed significance value. For these data the experimental conditions are not significantly correlated because $p > 0.05$.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Real Spider & Toupee	12	.545	.067

The final table tells us whether the difference between the mean in the two conditions was large enough to *not* be a chance result. First, the table tells us the mean difference between scores (this value is $\overline{D}$ in the dependent t-test equation) and represents the difference between the mean scores of each condition—e.g. $40 - 47 = -7$). The table also reports the standard deviation of the differences between the means and more importantly the standard error of the differences between subjects' scores in each condition. This standard error is the standard deviation of the distribution of differences between means from every pair of samples taken from a population. The test statistic, *t*, is calculated by dividing the mean of differences by the standard error of differences ($t = \frac{-7}{2.8311} = -2.47$). The value of *t* is assessed compared to known values and then a probability value is calculated. This probability value is the likelihood of the observed value of *t* happening by chance alone. This 2-tailed probability is very low ($p = 0.031$) and in fact it tells us that there is only a 3.1% chance that this value of *t* could have happened by chance alone. As social scientists we are prepared to accept that anything that has less than a 5% chance of occurring is statistically meaningful. Therefore, we can conclude that exposure to a real spider causes a significant increase in the level of measured anxiety in spider phobics ($p < 0.05$). The final thing that is provided by this output is a 95% confidence interval. Imagine we took 100 pairs of samples from a population and calculated the differences between sample means, then calculated the mean of these differences ($\overline{D}$). We would end up with 100 mean differences and this interval tells us the boundaries within which 95 of these 100 means would lie[1]. So, 95% of mean differences will lie between –13.23 and –0.77. The importance of this interval is that it does not contain zero (i.e. both limits are negative) because this tells us that in 95% of samples the mean difference will not be zero. If we took random samples from a population we would expect most of the differences between sample means to be zero. Therefore, if 95% of differences are not zero, we can be confident that the observed difference in sample means is due to the experimental manipulation.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Real Spider - Toupee	7.0000	9.8072	2.8311	.7688	13.2312	2.473	11	.031

[1] For those of you well up on the properties of the normal distribution, you should appreciate these limits represent the value of two standard deviations either side of the mean of the sampling distribution (in this example these values will be $-7 \pm (2 \times 2.8311)$). For those of you not familiar with the properties of the normal distribution read Rowntree (1981) *Statistics without tears*, chapters 4 & 5.

Homework: A psychologist was interested in then effects of music on aggression. One group of students listened to a tape of Billie (non-aggressive), and another group listened to a tape of Metallica (aggressive). A blow-up punch bag with a photo of either Billie or Metallica was placed in the room too. The dependent variable was the number of times the subjects hit the punch bag during the song.

Analyse these data (**Music.sav**) to find out whether music causes aggressive behaviour.