Questions and answers for Chapter 7

1. Open the data set. Can you use a test to see whether girls do better than boys at maths?

To look at this we would use maths grades as our dependent variable, and gender as our independent variable. Maths grades is continuous, and gender is nominal, so we can use a T-test to see whether there is a significant difference between boys and girls in maths grades. The output from the independent T-test should look something like this:

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
school grades maths	boy	304	75.5544	12.34581	.70808
	girl	271	76.4732	12.06939	.73316

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
school grades maths	Equal variances assumed	1.921	.166	-.900	573	.368	-.91880	1.02060	-2.92336	1.08577
	Equal variances not assumed			-.901	568.135	.368	-.91880	1.01927	-2.92079	1.08319

In the first box we can see that the mean grade for boys is 75.5, and for girls 76.5. So the girls' mean is somewhat higher. Is this difference significant though? To answer that question we need to look at the second output box. There we first need to look at Levene's test of Equality of variance. We can see that the significance of Levene's F-test is .166. This is larger than .05, so the difference in variance between boys and girls is not significant. That means that for the T-test we can assume that variances are equal, and look in the first row. We can see there that our significance level for the T-test is .368. This is more than .05, so the small difference we found between boys and girls in mean maths score is not significant. This means that we can't be sure that the small difference in this sample is not a coincidence. So we can't say there is likely to be a difference

between boys and girls in maths grades in the population.

2. How strong is the difference between the two groups' maths achievement?

   When we talk about the strength of a relationship, we have to look at the effect size. For the T-test, that means we have to calculate Cohen's d. There are a couple of steps in doing this. The first is to calculate the pooled standard deviation. This we do by adding the standard deviations for boys and girls (which can be found in the first output box) and dividing that by two:

   12.34+12.06=24.40

   24.4/2=12.2

   The second step is to calculate the difference between the means of the two groups (boys and girls):

   75.55-76.47=-0.92

   In the final step we divide the difference between the means by the pooled standard deviation:

   -0.92/12.2=-0.075.

   As this is less than -.1, we would say that this is a weak effect.

3. Can you compare boys and girls on the item 'school is boring' using the T-test? Explain.

The T-test works by comparing the means between two groups. Means are only relevant for continuous variables. 'School is boring' is an ordinal variable, so we can't use the T-test.

4. Can you compare boys and girls on the item 'school is boring' using the Chi Square test? Explain.

   The Chi Square test works by comparing actual values to those we would expect if there was no relationship between the two variables. This is based on a cross tabulation that shows the number of cases falling into each combination of the categories of two or more variables. The test is therefore suitable for variables with a limited number of categories. Gender has two categories (boy/girl) and school is boring has four (agree strongly/agree/disagree/disagree strongly). Therefore it is a suitable test to use here.

5. How strong is the difference between the two groups on this item?

   To answer this question we need to do a cross tabulation, calculate a Chi Square significance test (SPSS will of course do this for us if we tell it to by clicking the right box) and a measure of effect size. The measure of effect size we will use is Phi.

Our output should look something like this:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
school is boring * gender	888	99.9%	1	.1%	889	100.0%

school is boring * gender Crosstabulation

			gender		Total
			boy	girl	
school is boring	agree strongly	Count	77	44	121
		Expected Count	60.5	60.5	121.0
	agree	Count	105	128	233
		Expected Count	116.5	116.5	233.0
	disagree	Count	129	109	238
		Expected Count	119.0	119.0	238.0
	disagree strongly	Count	133	163	296
		Expected Count	148.0	148.0	296.0
Total		Count	444	444	888
		Expected Count	444.0	444.0	888.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.992[a]	3	.001
Likelihood Ratio	16.118	3	.001
Linear-by-Linear Association	5.449	1	.020
N of Valid Cases	888		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 60.50.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.134	.001
	Cramer's V	.134	.001
N of Valid Cases		888	

*In the table labelled 'school is boring*gender crosstabulation' we can find the expected values if there was no difference between boys and girls, and the actual values in the dataset. If there was no difference between boys and girls we would expect 60.5 boys and 60.5 girls (this number is of course a statistical artefact) to agree strongly that school is boring. In fact, 77 boys and 44 girls agreed strongly that school is boring. So boys are clearly more likely to agree strongly that school is boring than girls. However, girls are somewhat more likely to agree that school is boring than would be expected if there was no difference between boys and girls, and girls are also more likely to disagree. Boys are more likely to disagree strongly. Overall, then, boys appear to be more likely to give an extreme answer (strongly agree or strongly disagree) than girls.*

Is this difference statistically significant, or, in other words, can we be reasonably confident that this difference reflects a difference in the population rather than random sample fluctuation? To determine this we need to look at the Chi Square significance test. In the third box we can see that the significance level of Pearson's Chi Square is given as .001. This is a lot less than .05 and thus highly significant. We can therefore be reasonably confident that the difference in this sample represents a difference in the population.

Finally, we need to answer the question of how strong the difference is. To do that we use our effect size measure, Phi. This is .134, which suggests a modest difference between boys and girls on this variable.

6. Do you think it would be a good idea to stop using the Chi Square test and the T-test altogether, and use effect size indices instead? Explain.

As you know, Chi Square is a test of statistical significance. There is currently quite a debate about the use of significance tests in the statistics community. What pretty much everyone agrees on is that it is not sufficient to rely only on significance levels, as they are determined in part by sample size and therefore are a poor guide of the strength of a relationship. Therefore we should in all cases calculate an effect size measure as well as a significance test. Some researchers say that, in part because significance tests are so

often misinterpreted as effect size measures, and because of the arbitrary nature of the cut off points, it would be better to not use significance tests at all. Most researchers, however, feel that they still have a useful role to play alongside effect size measures.