Questions and answers for Chapter 8

1. You want to look at the relationship between pupils' responses to the item 'I think I'm good at maths' and their grades in maths. Which method do you use and why?

   To answer this question we need to consider what types of variables we are using. 'I think I'm good at maths' is an ordinal variable, while grades in maths is a continuous variable. The rule when we have two variables of a different level of measurement is that we need to use the method suitable for the variable at the lower level of measurement. In this case that is the ordinal variable 'I think I'm good at maths'. The measure we use to look at the relationship between two ordinal variables is the Spearman's Rho rank-order correlation coefficient, so we will use this to look at the relationship between an ordinal and a continuous variable as well.

2. Open the dataset. Is there a relationship between pupils' responses to the items 'I get good marks in maths' and 'I don't like the way I look'? How strong is the relationship?

Both these variables are ordinal, so we need to use the Spearman's Rho rank-order correlation coefficient. Doing this gives us the following output:

Correlations

			good marks in maths	don't like the way I look
Spearman's rho	good marks in maths	Correlation Coefficient	1.000	.067*
		Sig. (2-tailed)	.	.045
		N	885	884
	don't like the way I look	Correlation Coefficient	.067*	1.000
		Sig. (2-tailed)	.045	.
		N	884	885

*. Correlation is significant at the 0.05 level (2-tailed).

We can see that Spearman's Rho is .067. This is less than .1, so there is only a weak relationship between these two variables. The significance level is .045. This is less than .05 so the relationship is significant. So we can say that there is a weak but significant relationship, in that respondents who agree that they get good marks is maths are slightly more likely to disagree that they don't like the way they look (look at the value labels for this).

3. If there is a relationship between pupils' responses to the items 'I get good marks in maths' and 'I don't like the way I look', does that imply causation? Why, why not?

No. Correlation doesn't in itself imply causation. Remember that there were three conditions for causality: there is a relationship between a and b, a comes before b in time, and no third variable is the underlying cause of the relationship between a and b. Correlation only tells us about the first condition, not about the two others.

4. Open the dataset. Is there a relationship between age in months and grade point average? How strong is the relationship?

Age in months and grade point average are both continuous variables, so we can use Pearson's r correlation coefficient to look at this question.

Correlations

		age in months	grade point average
age in months	Pearson Correlation	1	-.274**
	Sig. (2-tailed)		.000
	N	885	883
grade point average	Pearson Correlation	-.274**	1
	Sig. (2-tailed)	.000	
	N	883	887

**. Correlation is significant at the 0.01 level (2-tailed).

In the output we can see that Pearson's r is -.274. So there is a modest relationship between the two variables. The significance level is .000, which means the relationship is highly significant (and therefore it is likely that there is a relationship between the two variables in the population as well as the sample). The sign is negative, which means that as age increases, grade point average decreases. So on average older pupils have a lower grade point average.

5. If there is a relationship between age in months and grade point average, does that imply causation? Why, why not?

   No, again we can't say that with any certainty. Think again of our three conditions. The first, that there is a relationship between the two variables, is demonstrated through correlation. The second, that a must come before b in time, is usually trickier, but not in this case. It is clear that age comes before grades (after all, doing a maths test doesn't affect ones biological age, though it may do so emotionally!). The final condition, that there are no third underlying variables causing the relationship is, however not demonstrated. One variable that we haven't measured may be particularly important here, and that is grade retention, widely used in the education system this dataset was collected in, as it means that the older pupils may be those that have been grade retained.

6. Above, I said that the correlation coefficient treats all relationships as linear, but not all are. Can you think of some examples of non-linear relationships that might occur in educational research?

   There are a wide range of non-linear relationships in educational research. One example is the relationship between teacher subject knowledge and pupil outcomes. Research generally finds that there is a positive relationship, i.e. more subject knowledge leads to better outcomes, but that this effect tails off, and has largely the following shape: