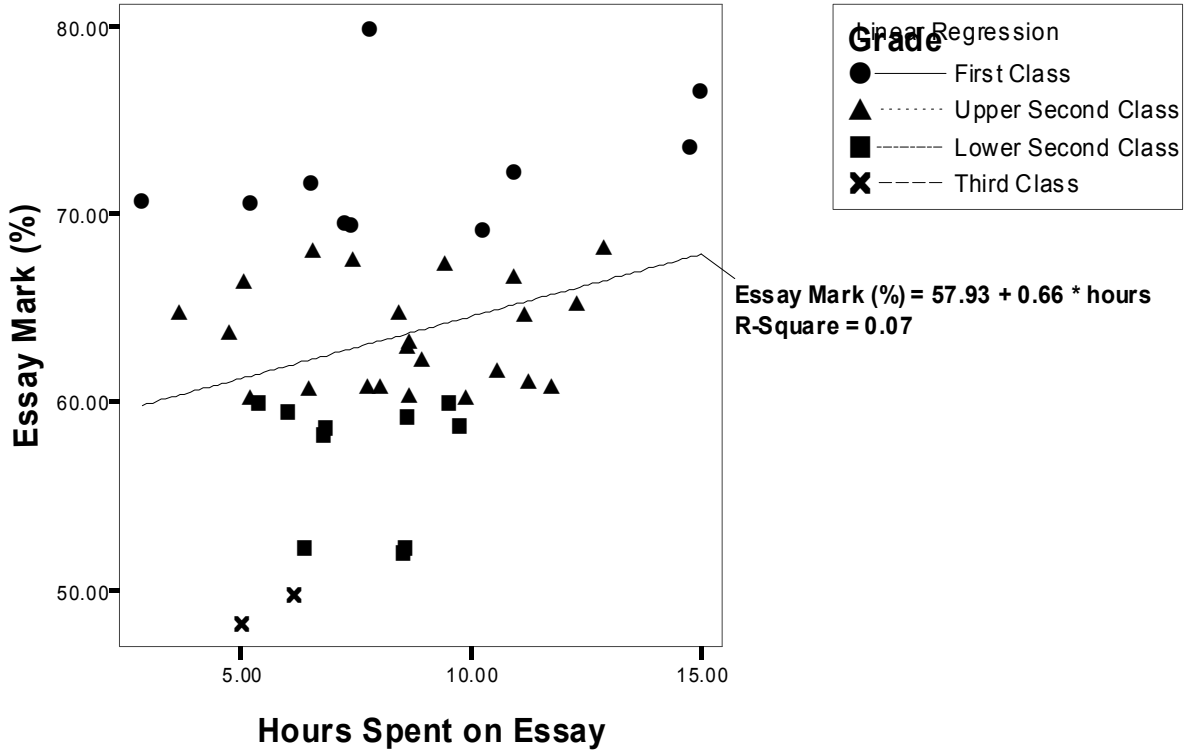
Chapter 4: Answers

Task 1

A student was interested in whether there was a positive relationship between the time spent doing an essay and the mark received. He got 45 of his friends and timed how ling they spent writing an essay (**hours**) and the percentage they got in the essay (**essay**). He also translated these grades into their degree classifications (**grade**): first, upper second, lower second and third class). Using the data are in the file **EssayMarks.sav** find out what the relationship was between the time spent doing an essay and the eventual mark in terms of percentage and degree class (draw a scatterplot too!).

We're interested in looking at the relationship between hours spent on an essay and the grade obtained. We could simply do a scatterplot of hours spent on the essay (*x*-axis) and essay mark (*y*-axis). I've also chosen to highlight the degree classification grades using different symbols (just place the variable **grades** in the *style* box). The resulting scatterplot should look like this:



Next, we should check whether the data are parametric using the *explore* menu (see chapter 3). The resulting table is as follows:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Essay Mark (%)	.111	45	.200*	.977	45	.493
Hours Spent on Essay	.091	45	.200*	.981	45	.662

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The K-S and Shapiro-Wilk statistics are both non-significant (*Sig.* is > .05 in all cases) for both variables which indicates that they are normally distributed. As such we can use Pearson's correlation coefficient. The result of which is:

Correlations

		Essay Mark (%)	Hours Spent on Essay
Essay Mark (%)	Pearson Correlation	1	.267*
	Sig. (1-tailed)	.	.038
	N	45	45
Hours Spent on Essay	Pearson Correlation	.267*	1
	Sig. (1-tailed)	.038	.
	N	45	45

*. Correlation is significant at the 0.05 level (1-tailed).

I chose a 1-tailed test because a specific prediction was made: there would be a positive relationship, that is, the more time you spend on your essay, the better mark you'll get. This hypothesis is supported because Pearson's r = .27 (a medium effect size), p < .05, is significant.

The second part of the question asks us to do the same analysis but when the percentages are recoded into degree classifications. The degree classifications are ordinal data (not interval): they are ordered categories, so we shouldn't use Pearson's test statistic, but Spearman's and Kendall's ones instead:

Correlations

			Hours Spent on Essay	Grade
Kendall's tau_b	Hours Spent on Essay	Correlation Coefficient	1.000	-.158
		Sig. (1-tailed)	.	.089
		N	45	45
	Grade	Correlation Coefficient	-.158	1.000
		Sig. (1-tailed)	.089	.
		N	45	45
Spearman's rho	Hours Spent on Essay	Correlation Coefficient	1.000	-.193
		Sig. (1-tailed)	.	.102
		N	45	45
	Grade	Correlation Coefficient	-.193	1.000
		Sig. (1-tailed)	.102	.
		N	45	45

In both cases the correlation is non-significant. There was no significant relationship between degree grade classification for an essay and the time spent doing it, ρ = −.19, *ns*, and τ = −.16, *ns*. Note that the direction of the relationship has reversed. This has happened because the essay marks were recoded as 1 (first), 2 (upper second), 3 (lower second),and 4 (third) and so high grades were represented by low numbers!

This illustrates one of the benefits of NOT taking continuous data (like percentages) and transforming them into categorical data: when you do, you lose information and often statistical power!

Task 2

Using the **ChickFlick.sav** data from Chapter 3, is there a relationship between gender and arousal? Using the same data, is there a relationship between the film watched and arousal?

Now, both gender and the film watched are categorical variables with two categories. Therefore, we need to look at this relationship using a point-biserial correlation. The resulting tables are as follows:

Correlations

		Gender	Arousal
Gender	Pearson Correlation	1	-.180
	Sig. (2-tailed)	.	.266
	N	40	40
Arousal	Pearson Correlation	-.180	1
	Sig. (2-tailed)	.266	.
	N	40	40

Correlations

		Film	Arousal
Film	Pearson Correlation	1	.638**
	Sig. (2-tailed)	.	.000
	N	40	40
Arousal	Pearson Correlation	.638**	1
	Sig. (2-tailed)	.000	.
	N	40	40

**. Correlation is significant at the 0.01 level

In both cases I used a 2-tailed test because no prediction was made. As you can see there was no significant relationship between gender and arousal, r_{pb} = −.18, *ns*. However, there was a significant relationship between the film watched and arousal, r_{pb} = −.64, $p < .001$. Looking at how the groups were coded, you should see that Bridget Jones' Diary had a code of 1, and Momento had a code of 2, therefore, this result reflects the fact that as film goes up (changes from 1 to 2) arousal goes up. Put another way, as the film changes from Bridget Jones' Diary to Momento, arousal increases. So, Momento gave rise to the greater arousal levels.