Chapter 3: Answers

Task 1

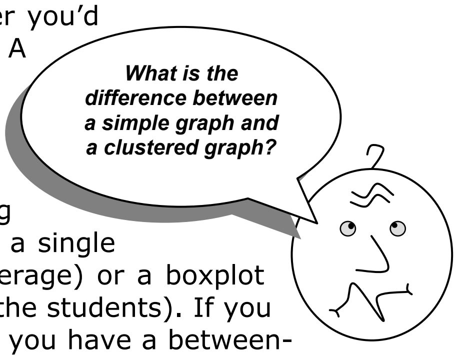
Using the data from Chapter 2 plot some graphs of the mean number of friends, alcohol consumption, income and neuroticism of students and lecturers. Which have the most friends, drink the most, earn the most and are the most neurotic?

We're interested in looking at the differences between lecturers and students. We took a random sample of five lecturers and five psychology students from the University of Sussex and then measured several variables: how many friends they had, their weekly alcohol consumption (in units), their yearly income (in pounds) and how neurotic they were (higher score is more neurotic). These data are in Chapter 2 and reproduced below. You should enter them into the SPSS data editor using what you learnt in Chapter 2. You should've already entered and saved these data, so retrieve the file.

Table 1: Data for differences between students and lecturers

Type of Person	*No. of Friends*	*Alcohol Consumption*	*Income (p.a.)*	*Neuroticism*
Lecturer	*5*	*10*	*20000*	*10*
Lecturer	*2*	*15*	*40000*	*17*
Lecturer	*0*	*20*	*35000*	*14*
Lecturer	*4*	*5*	*22000*	*13*
Lecturer	*1*	*30*	*50000*	*21*
Student	*10*	*25*	*5000*	*7*
Student	*12*	*20*	*100*	*13*
Student	*15*	*16*	*3000*	*9*
Student	*12*	*17*	*10000*	*14*
Student	*17*	*18*	*10*	*13*

Having entered these data we can look at trends by using graphs. To draw a graph, simply click on the word describing the graph that you want to plot. In most cases you will be presented with a provisional dialog box asking you whether you'd like to plot a *simple chart* or a *clustered* one (see Figure). A simple chart is one in which you plot one graph element per group or variable. For example we might want to plot the average number of friends for lecturers and students. As such, we want one bar representing the average number of friends that lecturers had and one representing the average number that students had. We could also plot a single line (connecting the lecturers' average to the students' average) or a boxplot (one representing the lecturers' data and one representing the students). If you want to plot one bar for different groups (this is used when you have a between-group design), then you should select *Summaries for groups of cases*. If you have only one group and you want to plot a graph of several dependent variables (a repeated measures design) then select *Summaries of separate variables*. An example of this would be if we ignored whether the person was a lecturer or student and just wanted to plot the average number of friends and the average neuroticism score on the same graph.



A clustered chart is one in which each group or category of people has several chart elements. These graphs can be useful if you want to plot two independent variables. For example, if we had noted the gender of each student and lecturer we would have two independent variables (gender and job) and one dependent variable (number of friends). Therefore, we could use a clustered plot to display the average number of friends for lecturers and students and have separate bars (or lines) representing males and females. In this latter case, in which both variables were measured using a between-group design, the *Summaries for groups of cases* option should be used. Alternatively, you can plot values of several groups along several variables. Imagine we wanted to plot the average number of friends and the average neuroticism score and split these scores according to whether the person was a lecturer or student. We want to plot two bars for the lecturers (one representing the number of friends and one representing neuroticism) and two for the students. To plot this graph we should choose a clustered chart but ask for *Summaries of separate variables.*

When a type of graph has been selected (simple or clustered) you need to click on Define to move to the next dialog box. On many occasions you will see the term *Category axis*, and this refers to the *X*-axis (horizontal). This axis usually requires a grouping variable (in these examples I have used *type of person*). Variables can be selected by clicking on them in the variable list (left-hand side of dialog box) and moving them to the appropriate space by using the ▸ button. In the case of bar charts, you can make the bars represent many things (number of cases etc.), but on the vast majority of occasions you will want them to represent the mean value and so you should select *Other summary function* and then enter a variable. The default function is the mean, but clicking on Change Summary... enables this default to be changed. Once the graphs options have been selected click on OK and the graph will appear in the output viewer. These charts can then be edited by double-clicking on them in the viewer. This action produces a new window (called the *chart editor*) in which you can change just about any property of the graph by double-clicking with the mouse (have a play around with some of the functions!).

Try putting some of these principles into practice using our lecturer-student data. Figure 1 shows how to create a bar chart and boxplot of two of the variables measured for the lecturers. Follow these options and see whether you can re-create these graphs (remember that you can edit them to add bar labels and change the colours). A bar chart of the means is a useful way to see the pattern of results (i.e. which group got the highest scores). In Figure 1 the graph shows us at a glance that students, on average, have more friends than their lecturers. Boxplots tell us a little bit more. For one thing the whiskers on the plot (the lines that stick out of the top and bottom) give an indicator of the spread of scores. More important, unusual cases can be identified (outliers) because they are displayed as a dot outside of the main range of scores. In Figure, the boxplot displayed has a single outlier who is represented by the dot above the graph. This person is a student who drank rather more than the other students. The dark line also shows the median score, so we can tell that the median amount of units drunk was higher for students than lecturers. Try plotting graphs of some of the other variables.

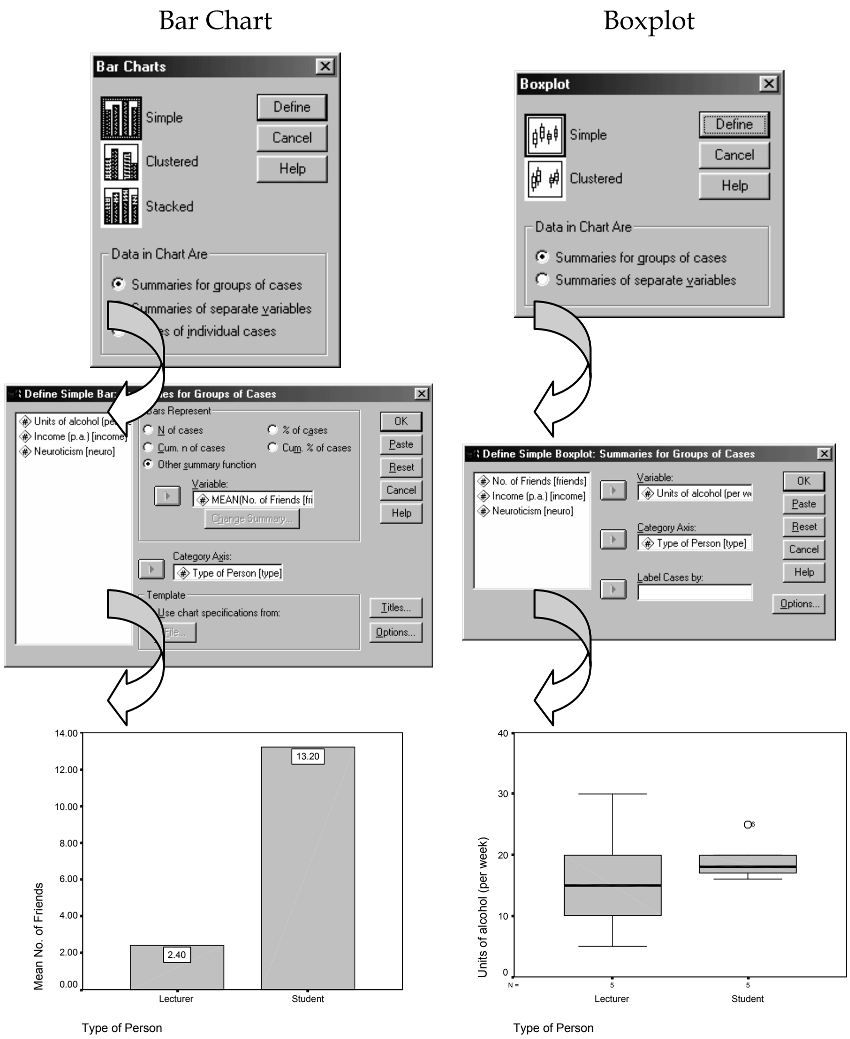


Figure 1: Plotting graphs on SPSS

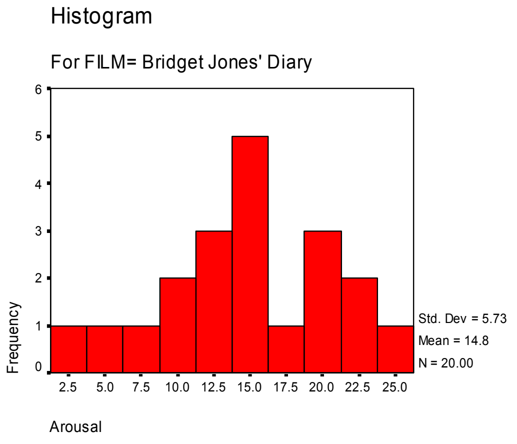
Task 2

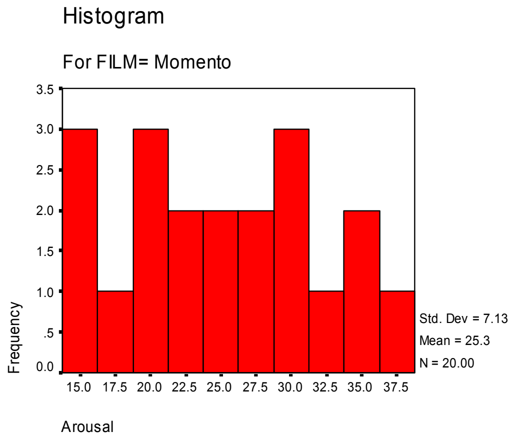
Using the **ChickFlick.sav** data, check the distributions for the two films (ignore gender): are they normally distributed.

The output you should get look like those reproduced below (I used the *Explore* function described in Chapter 3). The skewness statistics gives rise to a *z*-score of −0.378/0.512 = 0.74 for Bridget Jones' diary, and 0.04/0.512 = 0.08 for momento. These show no significant skewness. The K-S tests show no significant deviation from normality and the histogram for Bridget jones' diary even looks normal. The histogram for Momento is less normal, but the rest of the evidence gives us no reason to suppose it isn't.

Descriptives

	Film			Statistic	Std. Error
Arousal	Bridget Jones' Diary	Mean		14.8000	1.28062
		95% Confidence Interval for Mean	Lower Bound	12.1196	
			Upper Bound	17.4804	
		5% Trimmed Mean		14.9444	
		Median		15.0000	
		Variance		32.800	
		Std. Deviation		5.72713	
		Minimum		3.00	
		Maximum		24.00	
		Range		21.00	
		Interquartile Range		7.5000	
		Skewness		-.378	.512
		Kurtosis		-.254	.992
	Momento	Mean		25.2500	1.59419
		95% Confidence Interval for Mean	Lower Bound	21.9133	
			Upper Bound	28.5867	
		5% Trimmed Mean		25.2222	
		Median		24.5000	
		Variance		50.829	
		Std. Deviation		7.12944	
		Minimum		14.00	
		Maximum		37.00	
		Range		23.00	
		Interquartile Range		10.7500	
		Skewness		.040	.512
		Kurtosis		-1.024	.992





Tests of Normality

	Film	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Arousal	Bridget Jones' Diary	.127	20	.200*	.972	20	.788
	Momento	.097	20	.200*	.960	20	.552

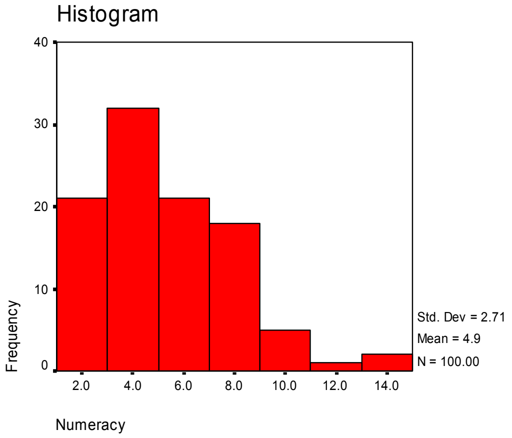
*. This is a lower bound of the true significance.

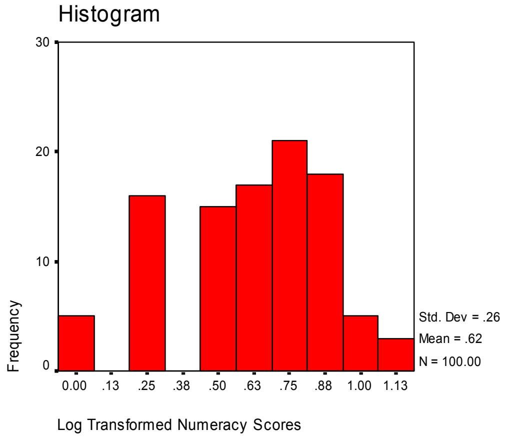
a. Lilliefors Significance Correction

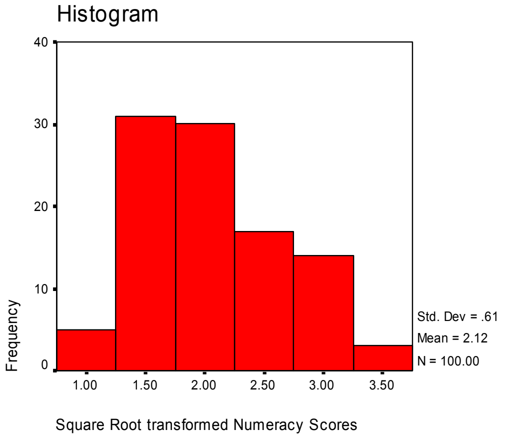
Task 3

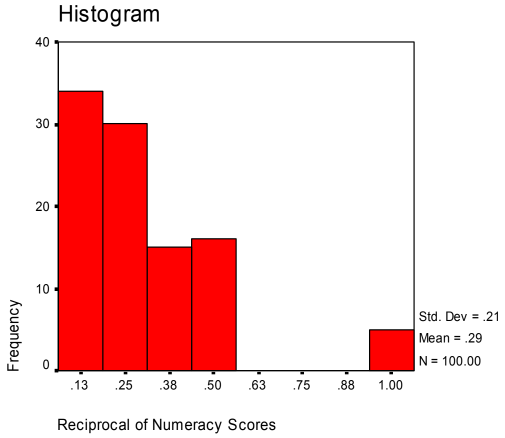
Using the **SPSSExam.sav** data, remember that numeracy scores appear positively skewed. Transform these data using one of the transformations described in this chapter: do the data become normal?

These are the original histograms and those of the transformed scores:









None of these histograms appear to be normal. Below is the table of results from the K-S test, all of which are significant. The only conclusion is that although the square root transformation does the best job of normalizing the data, none of these transformations actually works!

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Numeracy	.153	100	.000	.924	100	.000
Log Transformed Numeracy Scores	.120	100	.001	.959	100	.003
Square Root transformed Numeracy Scores	.108	100	.006	.970	100	.020
Reciprocal of Numeracy Scores	.223	100	.000	.763	100	.000

a. Lilliefors Significance Correction