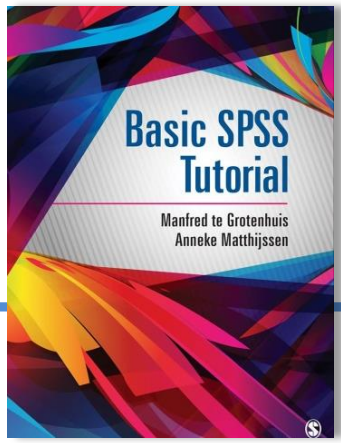
Basic SPSS Tutorial: chapter 4 Descriptive Statistics, extra assignments



This is a third set of extra assignments (1-9) related to descriptive statistics in *Basic SPSS Tutorial*. All computer related operations are placed in a blue-shaded box with the 💻 symbol. References to Basic SPSS Tutorial are indicated with BST.

The exercises relate to statistical relationships between interval/ratio variables. These relationships can be presented in a graphical way, like a scatter plot but also with measures of association like Pearson's correlation. Again, we use the measures of association for descriptive purposes only here. The data set in this series are from research on physical and mental health in the Netherlands in 2000.

💻
Download the data set BMI2000.SAV from the web page: http://study.sagepub.com/basicspss.
Start SPSS and open BMI2000.SAV (BST: section 2.2).
Open also a text file in the program Word or any other word processor where you store your answers to the questions below.

In BMI2000, you will find the variables *BMI* (that is the *Body Mass Index*, measured in kilograms per square meter) and *Education* (measured in years). A researcher wants to know whether BMI is to some extent determined by educational attainment. The idea is that BMI will tend to be lower for people with more years of education. The rationale behind it is that higher educated people have gained more insight in the determinants of overweight and will adapt their lifestyles accordingly.

1. Which variable is the dependent (y) variable and which variable serves as a predictor (x) variable?

💻
Create a scatter plot (BST: section 4.5, mind that x variables are always placed on the x-axis). In BST it is shown that scatter plots may not be that instructive. Therefore, create also a correct line graph for these two variables (BST: section 4.5).

2. What is the direction of the relationship: is it positive or negative? Also indicate what this means for the research question: does BMI increase or decrease when years of education increase?

💻
Paste the scatter plot into your text file (BST: section 4.7).
Most word processing programs like Word have a facility to include lines in the text file. Try to draw a (regression) line in the scatter plot and in the line graph that best fits the relationship.

The line you drew can be formulated as y = a + b * x, where a = the intercept (the average estimated score on y when x = 0) and b is the estimated change in y when x has an 1-unit increase (BST: section 5.8)

3. What are (roughly) a and b according to your regression line?

Calculate a and b with SPSS, use ordinary least square (OLS) linear regression analysis (see BST: section 5.8.1).

4. What exactly are a and b according to the statistical output?
5. What meaning has the intercept (a) in your analysis? Do you think this is a relevant outcome or not? Please explain why or why not.
6. What meaning has b in terms of BMI and years of education?
7. Take the values for a and b and calculate the estimated BMI for someone who has 6 years of education and for someone who has 13 years of education. Mind that the regression equation is BMI = a + b * education. How large is the estimated BMI for someone who has 6 / 13 years of education? See BST: p. 96 for calculation)
8. Take from the data set two respondents with 6 and 13 years of education respectively and show that the observed BMI and estimated BMI is not necessarily equal.

Calculate Pearson's correlation coefficient between *years of education* and *BMI* (BST: section 5.7)

9. What is the value for Pearson's correlation? Verify that this equal to the Beta in the OLS regression output (see BST: Table 5.9, lower part, 4th column and of course your own output)