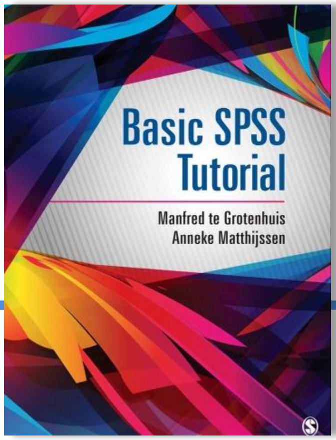
Basic SPSS Tutorial: chapter 4 Descriptive Statistics, extra assignments



This is a fifth and last set of extra assignments (1-8) 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.

We end the extra assignments on descriptive statistics with scaling (briefly mentioned in BST: section 3.3 and added some extra assignments that cover important parts of descriptive statistics.

Scales are summations of the scores on several variables (often called ‘items’). This summation is preferred over the use of separate items because a scale normally is a more reliable measurement than one single item. Sometimes scales consist of items from which it is clear that they have a lot in common. For instance a scale measuring crimes may be a summation of the number of pickpockets and the number of burglaries. In social sciences, however, items are often supposed to cluster together (it is assumed that all items measure more or less the same underlying subject) although it is unsure whether this is empirically the case. One way to find out is to look at correlations: items that correlate highly with each other have a lot in common and this is a necessary condition for measuring the same subject. Of course it is no sufficient condition: items may correlate highly and at the same time may be not related to the theoretical subject (this is the difference between *reliability* (high correlations between items)) and *validity* (do we measure what we want to measure). Besides looking at correlations there are more advanced ways to ascertain whether items cluster like *reliability analysis* and *principal factor analysis*.

💻

Download the data set SCALING.SAV from the web page: http://study.sagepub.com/basicspss.
Start SPSS and open SCALING.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.
Create a correlation matrix (see BST: section 5.7) for all 6 items that are supposed to measure Christian belief. The 6 items are named *v0149 v0150 v0156 v0159 v0166 v0180* respectively.

Note: strictly speaking these items are ordinal variables because the distances between subsequent categories are most likely not equal. In daily practice equal distances are assumed because mostly the differences are very small. The items therefore are considered to be interval variables and one may calculate Pearson’s correlations. Of course one can also calculate Spearman’s rank correlation (assuming ordinal measurement level) to check whether this leads to different outcomes.

1. Which of the 6 items correlates the least with the other 5?

💻

Create a scale (name: ‘*Christian_belief*’) with all 5! items that correlate highly with each other (BST: section 3.3) A scale constructed this way is called a Likert scale (see BST: p. 37).

To check whether this newly created variable is likely to be measure Christian beliefs, researchers often calculate its correlation with another variable which is indisputable connected to Christian belief. (One of these variables is *Church_attendance*. Check with both graphical and numerical tools whether the variable Christian_belief you created indeed correlates with church attendance (check direction and strength of the relationship). Copy all relevant information into your text file. See for all instruction the next box:

Create a line graph with *Church_attendance* (on the x axis) and *Christian_belief* on the y-axis), see BST: section 4.5.
Calculate Pearson's correlation between *Church_attendance* and *Christian_belief* (BST: section 5.7).

2. Argue why or why not the scale *Christian_belief* is a valid measurement of Christian belief (use the outcomes from all previous analyses with church attendance).

Someone argues that half of the Dutch population beliefs in a Christian God (i.e., they score 20 or higher on the variable *Christian_belief*).

Check with measurements of central tendency (BST: section 4.2) whether this might be true.

3. Explain why or why not it is plausible that half of the nation beliefs in a Christian God.

Someone argues that on average women believe more strongly in God compared to men.

Calculate for women and men the average score on Christian_belief separately (BST: section 3.5 and section 4.2. Copy the outcomes into your text file (BST: section 4.7).

4. Based on your statistical outcomes, it is likely that women believe more strongly in a traditional God than men? Why or why not?

Describe with the best suited graphical and numerical instruments the statistical relationship between the variables *Age* and *Christian_belief*. You may assume interval scale for both variables Copy all outcomes into your text file.

5. Based on the outcomes you just found, how would you describe the relationship between age and Christian beliefs (try to avoid technical/statistical terms).

> Run a linear regression analysis (BST: section 5.8.1) with *Christian_belief* as dependent variable and *Age* as independent variable.

6. Take the regression equation and compare a respondent who is 10 year older than another respondent. How much do these two respondents differ with regard to Christian belief?

> Create a new variable *Age2* which measures age in decennia (10 years). So type age2 = age / 10 in the dialog window 'compute variable' (BST: section 3.3 / Figure 3.3). Check whether the mean for *Age* is ten times larger than for *Age2*.
> Run a linear regression analysis with *Christian_belief* as dependent variable and *Age2* as independent variable (BST: section 5.8.1).

7. How large is the b-coefficient for *Age2*? Please explain the difference with the b-coefficient for *Age*.

Note: the magnitude of any b coefficient in linear regression analysis depends upon the scaling of the x variable as question 7 showed. Therefore, it is common practice to use beta coefficients instead when comparing the effects of x variables in regression analysis. This beta-coefficient is based on a standardized y variable and a standardized x variable (a standardized variable has mean=0 and standard deviation=1). For instance a beta of 0.5 means that y will increase with 0.5 standard deviation (of y) when x increases with exactly 1 standard deviation (of x). Because all betas in a regression model have this interpretation they are suited for comparison because the scaling is equal for all x variables.