


This is a fourth set of extra assignments (1-7) related to inferential 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.

In the third series about descriptive statistics we addressed cross tables. Then, the emphasis was on the correct percentages and the correct measure of association. In the exercises below we start with some repetition from chapter 4 (descriptive statistics) and thereafter we introduce the inferential part.



Download the data set WORK.SAV from the web page: <http://study.sagepub.com/basicspss>. Start SPSS and open WORK.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 cross table with the variables *Work* and *Sex* (see BST: section 4.6). Please make sure that the inner cells contain absolute counts + the correct percentages. The research question is about the inequalities between men and women (i.e., we like to compare the relative share of part time workers among men and among women). Paste your cross table into your text file see (BST: section 4.7).

1. What can be conclude from the cross table; is there a difference between men and women with regard to labor participation? Please explain which information let to your answer.

One may find in a cross table that indeed the relative share of full time workers among women is lower, which indicates an inequality. However that conclusion is based on a sample of the working population, so it remains to be seen whether this conclusion is valid for the nation as a whole. To get a more definite answer, we need a statistical test for the difference in percentages. In BST, section 5.2 we argue that the chi-square test is well suited.



Let SPSS compute the chi-square value in the cross table you created (BST: section 5.2, Figure 5.1).

- Note that a Chi-square test is always double sided in a 2 by 2 table. This is not correct here, because we like to test a directional hypothesis (the share of full timers is *lower* among women compared to men). How do we turn the chi-square test into a one-tailed test? (BST: p.81, top).
- How large is chi-square and is it significant at  $\alpha = .05$  (5%), one-tailed test?
- What is your conclusion about the gender differences in full time work?

The next question is about the relationship between educational level and church attendance (the variables can also be found in the data file WORK.SAV as well). From secularization theory the hypothesis is derived that the higher the educational level (x), the less one will attend church (y).



Create a cross table for the variables *Education* and *Church Attendance* (BST: section 5.2). Compute relevant percentages + the correct measure of association and test whether the relationship is significant. Paste the cross table together with the measure of association in your text file (BST: section 4.7).

- What is the outcome of the test with  $\alpha = .05$  (one-tailed) . Please explain which information led to your answer.

In the next three questions we introduce multivariate analysis. A bivariate relationship is analyzed and then we take into account a third (confounding) variable). We first like to test whether there is a positive relationship between income and the number of children



Create a cross table with the dichotomous variables *Income\_dummy* (0=<1500 Euros, 1=>1500 Euros) and *Child\_dummy* (0=no children or 1 child, 1=more than 1 child). Take *Kendall's tau b* and check whether its positive value differs significantly from zero (level of significance = .05, one-tailed).

- What is the outcome of the test; do people with a high income on average have more children than people with a relatively low income? Please explain.

Someone argues that the relationship between number of children and income is spurious (i.e, not causal) and is caused by age: the older one is the higher both the number of children and their income!



Check whether the relationship between *Income\_dummy* and *Child\_dummy* is the result of an aging process. Create a cross table with the dichotomous variables *Income\_dummy* and *Child\_dummy* for all people aged 18-30 and the same cross table for people aged >30jaar (you may use *Age\_dummy* in the data set as the 'layer' variable, see BST: p. 66, Figure 4.16). Test whether both *Kendall's tau b's* are still positive and reach significance.

- What is the outcome of the two tests ( $\alpha = .05$ , one-tailed)? Is the relationship spurious? Please explain.