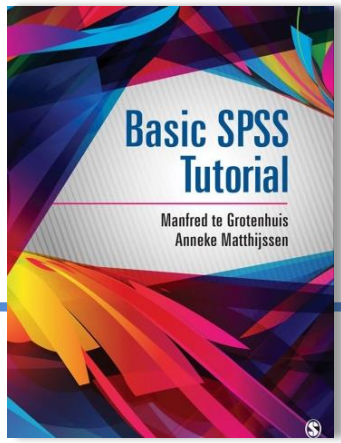
Basic SPSS Tutorial: chapter 5 Inferential Statistics, extra assignments



This is a fifth and last 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.

For obvious reasons older people have on average more children than younger people. This age-effect can be modeled in a bivariate regression equation: *Number of children* = a + b * *Age* + e, where b = positive.
One may wonder however whether this positive relationship between age and number of children is fully causal. It is a well-known fact that in Western societies youngsters have a quite different view on having children compared to older people (i.e., younger people settle for 1 or 2 children more often than older people). This is most likely not an aging effect but a 'cohort-effect' (i.e., people are born in different periods, and these periods shape people's view to some extent). If we take into account these cohort differences then the effect of aging might be reduced, because older people come from an era where having children had a different meaning. There is another cohort-effect that we have to take into consideration: older people have on average a lower educational level compared to younger people because the former are from an era in which the educational system was not as advanced as nowadays. Furthermore, higher educated people have less children compared to the lower educated. So, before we say something about a real aging effect we best take into account education and the attitude towards family and having children. This calls for a multiple regression analysis.

💻

Download the data set AGECHILDREN.SAV from http://study.sagepub.com/basicspss.
Start SPSS and open AGECHILDREN.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.

💻

Estimate the effect of *Age* on the *Number of children* in both a bivariate regression analyses and in a multivariate regression model taking into account the effects of educational level and the attitude towards children (cf. BST: section 5.8).
To take into account both confounding variables add the variables ATTITUDE and the dummy variables *LVS LSS, SS, Olevels, Alevels, College,* and *University* (*Elementary School* serves a reference) to the bivariate analysis: *Number of Children* = a + b * *Age*.

1. Is it true that the effect of age is lower after taking into account both confounding variables? Please explain.

2. Are the effects of educational level and the attitude toward having children as expected? Please explain. **Note**: estimates for all educational levels are deviations from Elementary School, a negative estimate means that respondents within that educational level have less children than people with elementary school only (the educational rank order is: *LVS* (low), *LSS, SS, Olevels, Alevels, College, and finally University* as highest level.

3. Test whether the confounding variables have significant effects at α = .05. What is your conclusion? Please explain.

The last section is related to the economic 'gender-gap'. Like in other countries, in the Netherlands several movements have tried to increase educational levels among women. Yet, in the Netherlands educational levels are lower than among men. Also the income among women is substantial lower compared to men (Dutch women earn on average 17.7% less than Dutch males).[1] In this assignment we will investigate the relationship between sex and income. We will look for one explanation: women may get less paid because they have lower educational levels compared to men.

One may find in a cross table that indeed the income workers among women is lower compared to men, which indicates 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 all working people. To get a more definite answer, we need a statistical test for the difference in percentages.

Create two separate cross tables with *Income* as the dependent (row) variable and subsequently *Sex* and *Education* as column variables (BST: section 4.6). Use the proper measurement of association and test for significance with α = .05. (BST: section 5.2). Paste these two tables into your text file (BST: section 4.7).

4. What is your conclusion, do *Sex* and *Education* relate to income in the assumed way? Please explain why/why not.

The relationship between gender and income might be caused by differences in educational levels (women hold on average lower educational levels)

Create a three way cross table to test whether *Education* is the explanation for the found relationship between *gender* and *income* (use *Education* as the 'Layer' variable, BST: Figure 4.16).

9. What is your conclusion, is it likely that the gender income gap is the result of differences in educational levels? Explain why/ why not.

When we consider income as a ratio variable, we could use a regression analysis as well to consider the effect of *Sex* on *Income* taking into account educational levels.

Use regression analyses (BST: section 5.8.2) and use *Income* as dependent variable and the dummy variables *LVS, LSS, SS, Olevels, Alevels, College,* and *University* (*Elementary School* serves a reference) + the variable *Sex*.

10. What is your conclusion, is it likely that the gender income gap is the result of differences educational levels? Explain why/ why not.

1. The International Trade Union Confederation (2009) *Gender (in)equality in the labor market: an overview of trends and developments* http://www.wageindicator.org/documents/publicationslist/publications-2009/GAP-09-EN.pdf