2 CORRELATION AND REGRESSION

2.07 Example contingency table

Suppose I conducted a study because I was interested in the question if people with a lower education are more inclined to believe in extraterrestrial life. I drew a sample of 261 respondents, which I selected randomly from the Dutch population, and asked them a couple of survey questions. Let me focus on 2 of those questions. One was formulated as the following statement: "The existence of extraterrestrial life is highly plausible." The respondent could choose between answering categories "Agree" and "Disagree". Another question concerned the level of education of the respondent. Respondents could choose from 3 categories: High, middle and low.

The main question I want to answer is if my two variables are correlated. The first thing I need to do is assess the level of measurement of my variables. As both variables are measured on an ordinal level, I will look at their contingency table. After all, a scatterplot, Pearson's r and OLS regression analysis are only allowed if variables are measured quantitatively. *This* is the contingency table. The categories of what I expect to be the independent variable – education – are in the columns, and the categories of what according to me is the dependent variable – belief in extraterrestrial life – are in the rows.

This table does not yet help me to determine if there is a correlation between the two variables. To assess if there is a relationship, we have to compute the relevant percentages. I will compute column percentages here. So, I proceed as follows. For every cell I compute the percentage of cases in that cell compared to the total number of cases in the column in which the cell is located. We have 25 cases here. The total number in its column is 50. This means that we get 25 divided by 50 multiplied with 100 equals 50 per cent. In this cell we have 26 respondents. The column total is 102, so we divide 26 by 102 and multiply again with 100. We get 25 per cent. We do that for all cells. *This* is the result.

What do these percentages show us? Are people with a lower education more inclined to believe in extraterrestrial life? If we focus on respondents with a low level of education, we see that only 12 per cent agrees with the statement that it is likely that there is extraterrestrial life. However, 63 per cent disagrees with this statement. If we look at respondents with a high level of education, we see the opposite. 50 percent agrees with the claim that it is likely that there is extraterrestrial life, whereas only 16 percent disagrees with this claim.

These percentages indicate that not the *lower* educated, but the *higher* educated are more likely to believe in extraterrestrial life! You can only discern that pattern when you work with percentages or proportions. So *always* compute percentages, and never work with the original numbers. Otherwise you will get strongly alien results!