Workshop Excel for data analysis

University of Applied Sciences Utrecht, Jonas Moons

Microsoft Excel is a powerful tool if you know how to deal with it. Excel is also the standard spreadsheet program within the business world. For these reasons, it is very useful to have more than just basic knowledge of Excel.

In this exercise we will work with a real data set. Respondents received a Fitbit activity tracker, which they could then try a couple of months. The research focused on predicting behavior (wearing the Fitbit). The data we have is exactly as I found it myself. To make it extra challenging: it is a French-language study. So, time to brush up your French.

We will work on the basis of *use cases* : a specific task that you want to perform and for which you need specific skills. For each use case, I will demonstrate the relevant Excel functions, after which you get to work with the file.

1. You want to get an idea of the file and the variables

Explore the file:

- Which (type of) variables are there? Which worksheets are there? What do these worksheets mean? How are they linked together?
- Explore the quantitative variables

Topics in the explanation:

- Basic Excel
  - CTRL-X, CTRL-C, CTRL-V, CTRL-Z, CTRL-Y , CTRL-A
  - Selecting cells with SHIFT and CTRL
  - Moving and inserting columns
  - Copying and pasting data (to multiple cells)
  - Filling cells by dragging down
- Sort and filter by column titles
- Block rows and columns
- Hide/show rows/columns
- Make rows/columns wider and narrower
- Sum, mean and count at a glance

2. You want to clean and categorize a text variable

In the AZ - BB columns people have entered their health apps. The way in which people wrote down their apps, however, differs a lot. You want to ensure consistent coding. You also want to create new variables, which indicate per app whether the user has installed them or not.

Assignment: remove all '-' in the three columns with CTRL + H. Create a new variable for 'Croix rouge'

Topics in the explanation:

- CTRL + F
- CTRL + H

- Quantitative coding of open answers

3. You want to create a graph

You want to create a new worksheet with a graph of the number of inhabitants per city included in the study (look up the numbers online).

Topics in the explanation:

- Graphs
- Adjusting graphs

4. You want to use Excel charts in Word

You want to use data from Excel in different ways in Word, both linked and not linked.

Topics in the explanation:

- How you can integrate Excel files into Word

5. You want to examine the relationship between variables, where at least one variable is qualitative (consists of categories, not numbers)

Explore how men and women differ from each other on the different variables.

Topics in the explanation:

- Pivot tables
- Percentages in pivot tables
- Pivot graphs

6. You want to calculate a new variable

You want to create a new variable with the BMI. You also want to classify this variable into categories: overweight or not. You only want numbers in your final file, not formulas.

Topics in the explanation:

- Formulas: basic, copying formulas
- Formulas: pasting values
- Formulas: the IF()/ALS() formula

7. You want to examine the relationship between two quantitative variables (consisting of numbers)

Investigate the relationship between 'qualité sommeil ' (quality of sleep) and 'level moral' (mood). Is it a linear relationship (r real line)? What is the formula?

Topics in the explanation:

- Scatter plot
- Trend line
- Linear regression

8. Manipulating text data

You want to create one column with a text field with all apps that the user has, separated by a comma.

Topics in the explanation:

- Text formulas
- CONCATENATE() / TEKST.SAMENVOEGEN()
- "text"

9. You want to link two data files on the basis of a common key variable

Linking two files is a task that you may encounter a lot. Suppose, for example, that we want to get the average number of steps in the survey data. You then have to work with a common 'key'. In this case, that is the column 'id'. This is quite complicated (a lot more complicated than in Python!) and we will only deal with it if there is time left.

See here for the explanation.