Assignment: How to use SPSS?

|  |  |
| --- | --- |
|  | **Unit 4:** The aim of this assignment is to get to know SPSS and practice some basic aspects of this statistical program. |
|  | 1. [SPSS for absolute beginners](https://player.vimeo.com/external/233471203.hd.mp4?s=26b072432d2da35b5fa2a17458564ee4c85a5c05&profile_id=175)2. [introduction SPSS](https://vimeo.com/user29453510/review/139201671/8ac05b25e8)3. [Changing variables in SPSS using recode](https://player.vimeo.com/external/233476822.hd.mp4?s=c5d2fc8a24cd864b286b463adc84c03c91fa5861&profile_id=175)4. [A data matrix in SPSS](https://player.vimeo.com/external/233456220.hd.mp4?s=c14a2ddecb77a0540a4279d968b0a6ac2409c94a&profile_id=175)5. [Frequency tables](https://vimeo.com/user29453510/review/139203937/a0eac91269) |  | Data matrix; Codebook; Statistical programs (SPSS/ R / Stata); SPSS: syntax file, commands, menu, output file, data matrix, variable overview, variable and value labels  |
|  | No additional readings |  | Form groups of 4 (max), face each other. |  | 90 minutes |
|  | Download and open the datafile “Unit 4 - Data.sav” from Blackboard.Read this assignment carefully and answer the questions. Bring either a print or digital version to the tutorial. |

1. SPSS is using three different types of files. Shortly explain the main purpose of each file type.

|  |  |
| --- | --- |
| **Type of file:** | **Main purpose:**  |
| data matrix (.sav) | *All data to be handled are collected* |
| output file (.spv) | *Used to communicate the data* |
| syntax file (.sps) | *Includes commands used to transform the data matrix into output* |

1. Suppose you work in a research team and your job is to analyse the data you have collected. After your job is done, you want to share your work with the team. Which data files would you choose to share with your team? Explain for every file type why you think it is important to share it.

|  |  |
| --- | --- |
| **Type of file:** | **Sharing with the team? Why (not)?**  |
| data matrix (.sav) | We may want the team to see the variables and what they mean so we may also send the data matrix. |
| output file (.spv) | We want to mainly to share the output data to make sense of what the data means.  |
| syntax file (.sps) | We may also like the team to see the syntax file so that the team sees the commands (questions) to the data matrix |

1. Data matrix. Suppose you have information about the age, gender, year at university, favourite course, for 100 students and you want to store this information in a data matrix. Would you choose the columns or the rows to refers to the students? And what about the four variables?

Columns: Variables

Rows: Individual cases

1. Open the downloaded data file “unit 4 - data.sav”. Explain the main difference between the dataview and the variable view.

Data view: We can see all the variables (in the columns) per individual case (in the rows), all the cells contain observations.

Variable view: We can see labels attached to the values of variables. We can see what the variable means, all characteristics of the variables.

1. The data file “Data.sav” is a data matrix with four variables.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Values in the data matrix** | ***Variable labels*** | ***Value labels*** |
| V1 | 0 and 1 | *Gender* | *0 = “Female”,* *1 = “Male”* |
| V2 | 19 to 29 | *Age* | *19 to 29* |
| V3 | 1 to 7 | *Year at University* |  |
| V4 | 1, 2 and 3 | *Favourite Course* | *1 = “Ethics”,* *2 = “Statistics”,* *3 = “Academical Writing”* |

Check the **data view.** Make sure you can understand where to find the variables and the units. How many students are in the datafile? 99 students.

1. Using the **variable view**,
	1. assign *variable labels* to the respective variable
	2. assign *value labels* to variables V1 and V4.
	3. specify the l*evel of measurement* of each variable
2. Using the **data view**. Add a 100th case to the data set using the following information:

The participant is a 26-year-old female student in her third year of studies who likes Statistics most.

1. SAVE your data file with a new name (for example by using \_2 after the original name).
2. Open a syntax file: “File -> New -> Syntax”. In this syntax file, type (or copy/paste) the following:

DESCRIPTIVES VARIABLES=V1 V2 V3 V4

 /STATISTICS= MIN MAX.

To run this Syntax, highlight it and click on “Run”, which is the green play button above the written syntax. Check the output file (which is automatically created).

1. **Use the menu**, go to “analyze -> descriptive statistics -> descriptives”, select the 4 variables, go to “options” and select ‘min’ and ‘max’ (unclick the others). Click ‘continue’ (which sends you back to ‘descriptives’) and **click on ‘paste’** (NOT ON OK). Check the **syntaxfile**. What do you see?

You now see the same command you just typed in, or copy pasted.

1. We are interested in the students who favor statistics (as compared to all other courses). We will therefore create a new, simplified version of V4. SPSS offers many options to do that. We will now use the **recode command**. Under the recode command there are two options: Into Same Variable and Into Different Variables. Into Same Variable replaces existing data with new values permanently. You cannot fix this later. The command Into Different Variables adds a new variable to the data set. In almost every situation, you want to use Into Different Variables. This was all explained in the micro lecture.

Create a new variable using *recode into different variables* of variable V4. Give this variable the name Statistics and the variable label ‘Statistics is favorite course'; value labels: 0 ‘no’ 1 ‘yes’.

1. Via the menu option, creating a new variable is a lot of clicking. Writing commands directly in the syntax is much quicker:

RECODE V4 (1=1) (2 thru 3=0) into Ethics.
VARIABLE LABELS Ethics ‘Student favors Ethics course’.
VALUE LABELS Ethics 0 ‘No’ 1 ‘Yes’.
EXECUTE.

Select the commands and run them.

1. Create the last dichotomous variable ‘Academicwriting’ via menu or syntax.
2. Request frequency tables for the variables V4 (the original variable) and the new variables Statistics, Ethics and Academicwriting. You will see four frequency tables in the output file.

Check if the creation of the new variables was successful.

Compare **the number of respondents** with an answer 1 in variable V4 with the number of respondents with a value 1 in the variable Ethics. Do the same for the other categories of V4. If they all match, the recoding was successful.

1. **ALWAYS save your syntax file, the data file and the output file under a different name**. With the syntax file you can run every command again to replicate the data file or output, in case you lost either of them due to computer malfunction or whatever other reason.