Lab 1: Introduction to SPSS

This lab will introduce you to how to:

- start up SPSS
- import new data into SPSS
- change the properties of the variables
- sort the data according to a variable
- generate basic frequency tables
- generate basic graphs

Starting up SPSS

- You will find SPSS under "All programs", in a folder titled "IBM SPSS Statistics". There might be multiple versions there, you want the version 22, and the one that says nothing else after it; it should have the greek letter Sigma on its icon.
- It might offer you some pop-up windows at the start, just exit out of them until you see an Excel-like spreadsheet view.

Importing new data into SPSS

- The next step would be to locate the file you want to load. In our case it is a file on the web, so:
  - Fire up your web browser.
  - start by going to the datasets list[1] for the class.
  - You will find there a dataset titled "cars". Right-click on it. Do NOT left-click.
  - Choose "Save link as" or something equivalent, and save the file somewhere that you could easily find.
- Now go back to SPSS. Open the File menu.
- Select the "Read Text Data" option. This is important to remember: For new data you typically need to do this "Read Text Data" process, and not use the normal "Open" menu.
- An "Open Data" window will have opened up.
  - Navigate to the folder where you saved your file.
  - If your file is not visible, select "All files" from the "Files of Type" pulldown menu.
  - Select the file and click "Open".
- This will start the "Text Import Wizard". It has 6 steps

[1] http://skiadas.github.io/AppliedStatsCourse/site/datasets.html

  - In step 1, look at the area at the bottom and make sure it does not look "weird". It should have rows of data. If it does not, then you have not downloaded your data properly. Make a note whether the first line of data is the variable names or not. Click Next.
  - In step 2, check the correct box where it asks you whether the top of the file has the variable names or not (in this case it should be "yes"). When you look at your data at the bottom, the variable names should no longer be there. Click Next.
  - In step 3 there is nothing for us to do. Click Next.
  - In step 4, the data should have been lined up in columns, and some delimiters chosen at the top left. These are used to separate the entries. Look through the data to see if anything seems out of place. You may find some "-1" entries, and that is OK, they represent missing values. Some times SPSS might have used more delimiters, that would result in columns being misaligned. In that case you may have to deselect those delimiters. Click Next.
  - In step 5, you are given the option to change the type of the columns variables. It is useful for data containing dates and times, but this dataset does not have any. We will return to this at a later lab.
  - In step 6 there is nothing to do. Finish the process.
- You should now be looking at your dataset in SPSS. go ahead and use "File -> Save" to save the dataset in the SPSS format. You can open those files in the future by double-clicking on them instead of having to do the Read Text Data steps.

Changing variable properties

- Use the "Data -> Define Variable Properties" menu.
- Select which variables you want to change.
  - Keep the Ctrl key down to select multiple variables.
  - Type Ctrl-A to select all variables. Do this in our example.
  - Click the arrow in the middle when ready. Then click Continue.
- The "Define Variable Properties" menu pops up.
- On the left is a list of the variables you have selected. The currently highlighted variable is the one you are working on at the moment. Select the Type variable for now.
- On the right is information about the current variable, and we shall edit that information now.
  - The first field says "Current Variable". Next to it it says "Label". This is the name for the variable as it would appear on a graph or table. If it is blank, the variable's shorter internal name is used. It is better to give more descriptive names here, including spaces. For instance put "Car Type" in.

- Right below is the “Measurement Level”. This is the classification of the variable as Scalar, Ordinal or Nominal. We will stick with Nominal in this case.
- The table a bit below that shows the different values together with counts of how many time each occured. If you see a value of -1, this usually indicates a missing value. If you have such a value, then you want to check the corresponding box in the “Missing” column.
- The “labels” column allows you to type more intuitive labels instead of the mysterious code numbers. Do this now for this variable. Use the arrows to move from one to the next. These are usually called “Value Labels”. Here are the labels for this variable:
  - 0 -> Standard
  - 1 -> Sports
  - 2 -> SUV
  - 3 -> Wagon
  - 4 -> Minivan
  - 5 -> Pickup
- Give more appropriate labels to all the variables, by choosing each variable on the left. Here are the labels:
  - Car Name
  - Car Type
  - All-Wheel-Drive
  - Front-Wheel-Drive
  - Manufacturer Price (USD)
  - Retail Price (USD)
  - Engine Size (liters)
  - Number of Cylinders
  - Horsepower
  - City Miles per Gallon
  - Highway Miles per Gallon
  - Weight (pounds)
  - Wheel Base (inches)
  - Length (inches)
  - Width (inches)
- Some variables have -1 as values. Mark those as missing values.
- Assign a proper “Measurement Level” to each variable. Most are correct, but at least one requires a change.

- When you are done click OK. An Output window may appear, simply click back at the data window.
- This is a good time to SAVE your dataset, so that you don’t have to repeat this process the next time you want to work with it.

Sorting the data

- Often times you want to have the rows sorted in a particular order. For instance suppose we want the most efficient cars, those with high CMpG, to be appear higher on the list.
- Right-click the name of the variable you want to use (CMpG in this case).
- Select "Sort Ascending" or "Sort Descending" depending on which is order you want (descending in our case).
- Win.
- What are the 3 most efficient cars in this list? Can you tell why?

Generating Frequency Tables

- Choose "Analyze -> Descriptive Statistics -> Frequencies"
- Make sure "Display frequency tables" is selected.
- Select the variables you want. In our case that would be Car Type and Cylinders.
- Click "OK". You shoul be seeing one table for each variable.
- Which Cylinder types are the most frequent?
- How many Sports car models were there? What percent of all the cars was that?

Generating Basic Graphs

- Choose "Graphs -> Chart Builder".
- Click OK if you get a pop-up. You should now be staring at the "Chart Builder" window.
- First choose a graph from the Gallery. We will start with Bar graphs. So Select the "Bar" option from the list at the bottom left, then double-click the first graph there, the one with the 3 plain bars. A model for it should show up on the top right area.
- There is an area marked for the "x axis". Drag there the variable that you want to correspond to the different columns. We will use Cylinders.
- The y-axis should now say Count. NEVER mess with the y-axis in a bar graph or a histogram. If you do, reset the graph and start from scratch.
- Click OK. You should see your graph in the viewer.
  - What conclusions can you draw about the number of cylinders on the different car models from this graph?
- You can find Pie charts in the "Pie/Polar" graph section. Try to make Pie charts for the number of cylinders and for the car types.
  - The Pie chart should be able to better give you the information that close to half of the car models have 6 cylinders. This information was not as easy to discern in the bar graph. DIFFERENT GRAPHS CONVEY DIFFERENT INFORMATION.