Lab 3: Investigating relationships between variables

Creating 100% stacked bar graphs

Used when both variables we want to compare are categorical.

- Start a bar graph as usual in "Graphs -> Chart Builder" and selecting "bar" at the bottom left.
- Choose the third option from the list of bar charts.
- Place the stacking variable at the "stack set color" spot, put the other variable on the x axis.
- Do NOT put anything on the y axis.
- Once you have created the graph, double-click to edit it and choose "Options -> Scale to 100%" from the menu.
- You can change the color of individual bars by selecting them.
- Remember that which variable goes where matters. You reach different conclusions.

Creating a Crosstabulation

A cross-tabulation is a two-dimensional frequency table for pairs of categorical variables. Each cell provides a count for the items that belong to the corresponding combination of values from the two variables.

- Go to "Analyze -> Descriptive Statistics -> Crosstabs"
- Select one variable for the rows and one for the columns.
- Under "Cells" you may include "Row" or "Column" percentages.
- Take some time to learn to read the table.

Adding a panel variable

Plots typically only show two variables. We can add a third categorical variable as a "panel variable", creating a different graph for each value.

- Start to make a plot (e.g. stacked bar chart) as usual.
- In the "Group/Point ID" tab, you can choose to set a "row panel variable" or "column panel variable".
- Place the variable in the new box that shows up.
- **Caution**: If you do not have enough cases to begin with, splitting them based on three different criteria can leave too few cases for each combination. For instance if the first variable has 4 different values, the second variable has 3 different values, and the third variable ha 2 different values, that is a total of $4 \times 3 \times 2 = 24$ different combinations, so our cases are placed into 24 different "buckets". If we have only 100 cases to start with, that means on average a mere 4 cases per bucket.

Splitting the file

Splitting the file allows us to produce output separated for each value of a categorical variable (e.g. separately for males/females).

- Menu: “Data -> Split File”. Do NOT choose “Split into Files”.
- Select “Compare Groups”.
- In the “Groups Based on” area, put the variable that you want to split by.
- Go back to the main view. Things will appear the same.
- Perform whatever computation you were intending to make (graph, frequency tables, descriptive statistics). You will get separate output for each value of the categorical variable.
- To revert to looking at all cases again, go back to the same menu and select “Analyze all cases, do not create groups”.

For practice you can try the “Analyze -> Descriptive Statistics -> Descriptives” menu in combination with a file-splitting.

Filter cases

Some times it can be useful to exclude some cases (e.g. outliers) from consideration. This option allows us to “temporarily hide” these cases.

- Menu: “Data -> Select Cases”.
- Choose “If condition is satisfied”.
- Click the “If” button. We get an “equation editor”.
- Write a formula that describes which cases we want to have **included**. All other cases will be excluded.
- When you go back to the data view, you will see some rows having their number crossed out.
- A new 0/1 variable will have been created, to indicate which cases are selected.
- Carry on whatever analysis you want, and only the selected rows will appear.
- To revert back to all cases, go back to “Data -> Select Cases” and choose “All cases”.
- To carry out the same selection, you can choose the otion “Use the variable”, instead of “if condition is satisfied”. Put the new variable that was created there.