A guide to useful tools for all websites

Weighting

Why use weights?

Many of the statistical tools we use, and particularly those attempting to infer from a specific sample to a wider population, are based on a random sample. However, may of the datasets we use are not completely random. This is for two main reasons:

1. Stratification was used. This means that the population was divided to different groups (for example rural and urban, or according to religious or racial categories). This is usually done when some of these groups are much smaller than others – for example there are less people who live in distanced rural locations. To sample enough people from these areas the sample would have to be very large. If these people are important for the analysis, sometimes the smaller groups will be over-sampled so that the sample will have enough of them, without becoming very large and expensive. This over-sampling can later be corrected by weighting.
2. Non-response. Despite an attempt for a random sample, some population groups (such as women and older people) tend to respond more to surveys. To correct for this, previous knowledge of the population can be used to give the proper weights to each group.

How to use weights?

In SPSS, from the Data menu, choose "Weight cases". In the new menu box choose "Weight cases by" and choose the appropriate variable. To stop the weights, go to the same menu and choose "Do not weight cases".

Missing values

Why use missing values?

Sometimes we don't want certain values of a variable in our analysis. These values are normally variations of "don't know", "refuse", "not relevant", and "not asked". Often, these categories get values such as 999, which will skew our analysis.

How to use missing values?

In SPSS in the variable view, hover over the "missing" cell for the variable you are interested in. 3 dots will appear, click on them and a menu will open which will allow you to define the missing values.

Filters

Why use filters?

Randomly sampled surveys often aim to represent the entire population in a country (with some restrictions, such as over 18). However, the population we are interested in might be a sub-section of that entire population. For example, we might be interested only in those living in rural areas. To do that, we would need to use a filter.

How to use filters?

In SPSS Data menu, choose "Select cases", click on "If condition is satisfied" and then of the "If" button to create your condition. It will usually look like this: Variable = Category, for example: Sex = 1. Then you can choose if you want to (1) filter out unselected cases - this option will still keep them in the dataset, and you could select "all cases" in the same window to delete the filter and keep doing analysis on the full sample, or (2) copy the selected cases to a new file, which you will then be able to save separately. Unless you have multiple copies of your file, the third option of deleting the unselected cases is not recommended.