General Directions for Analyzing Project Data

This is a general guide to help you get started in analyzing your data.

Work with your group members. Refer to your friendly neighborhood statistics instructor if you have any questions/problems.

Some steps may not apply in your particular case.

Preparations

- First off, you should document your work along the way in a Word document, so start a new Word document. Think of this as a **journal**, and not what you should turn in as a final report. It is *for your use only* to keep track of your progress. It is your notes, from which you can build your report. Start out by listing your variables and their types (quantitative, ordinal, nominal?).
- It often would happen, especially if you did a survey, that some of the answers are not in the form you expected, like for instance someone answering with the number 12 when you asked them for something between 1 and 10. For each of these, you need to decide what to do with it. For instance in the above case, you could turn the 12 into a 10, or you could consider it as an invalid answer and leave it blank. There are typically no right or wrong decisions here, but you have to make a decision. Make sure to document these decisions in your journal.
- Next, start SPSS and **import** your data set into SPSS format. If it is an Excel Spreadsheet, which it probably is, then you may need to save it as an "old Excel format" (2003) (but try without this step first). You will definitely need to CLOSE Excel. You cannot read any file into SPSS if it is open in Excel. Finally, in "Read text data", you can select the file type to be Excel.
- SAVE OFTEN
- Use the Variable Properties menu to change any variables whose type is incorrect. Remember, quantitative variables need to be in *scale*, ordered categorical variables in *ordinal*, and the unordered categorical variables in *nominal*.
- Make sure you understand what each variable measures. Give an appropriate **label** to each variable (Don't forget, this is what will appear in graphs for that variable, so it should be not overly long but descriptive enough, and should definitely include the units used).
- Check that the value labels for the categorical variables are what you would want them to be, and make any changes necessary. If you coded things with numbers, now is the time to give names to these numbers.
- If you had entered the variables with full words instead, you will want to use "Automatic recode" to convert them in a format that SPSS prefers. Before you do that, do frequency tables, and make sure you don't see the same value multiple times due to typos (for instance if you see the values 'Yes' and 'yes' showing up as different values, that needs to be corrected). You may also skip the "Automatic recode" step until it seems necessary.

Don't forget to SAVE OFTEN.

Explorations

- Did I mention you should SAVE OFTEN?
- The main point of the Explorations section is to **OWN your data**. You want to learn the inns and outs of all your variables, including what time they woke up and what they had for breakfast.
- For each variable, create appropriate graphs and numerical summaries to describe it. For categorical variables, frequency tables, bar graphs, pie charts etc would be appropriate. For quantitative variables, boxplots and histograms, means and standard deviations. Describe your observations in your journal. Along with whether they seem reasonable to you. Do the quantitative variables look like normally distributed, or do they deviate from normality? Are they skewed? Do they have outliers? etc
- You will need to prepare the graphs more carefully at a later day, but for now just copy and paste the appropriate graphs from the Viewer to your journal. Make sure to also **save the Viewer file** this time, to preserve all the work you did on the graphs, and make sure to document how you obtained the graphs, in case you need to get them again. The point is that you will need, at this point, to generate many more graphs than what you will include in the report and presentation, in order to understand your data. So at a later point in the project, once you have decided which graphs you want to include in your report, you can then customize them further. You cannot do this with the copies that are in Microsoft Word, so it is important to keep enough information around in SPSS to be able to recreate the graphs, if needed.
- At this point, you might decide that a particular variable should be grouped in 2-3 groups, instead of being kept as scalar. Or you might need to do some other kind of transformation. Do this now if necessary, using the "Recode" and "Compute Variable" menu options, to create these new variables. Don't forget to document this step, as with previous steps.
- Once you have gotten a feeling about how your variables are distributed individually, it is time to consider the **interactions** between variables. For instance with the number example above, suppose another variable is male/female. One of the questions would be: Do the numbers for males and females differ? Again graphs and numerical summaries can help with that.
- At this point you want to start thinking about the key questions you wanted to have answered in this report. Typically, those questions can probably be expressed as asking how one of the variables depends on another one, how they are associated. Using the graphs and numerical summaries established in the previous part, try to answer these questions (and document both the questions and their answers). One of the things you should look out for is whether some *lurking variable* could be influencing what you are seeing, by being confounded with the other variables or by explaining both variables (Recall our discussions about *common response* and *confounding* in the "Correlation and Causation" chapter from the book).
- Often you will need to examine the interaction between two variables, subject to the levels of a third variable. For instance, is the relation between weight and

height different for males and females? This third variable can appear for instance as a grouping variable in a scatterplot, or as a panel variable in a boxplot. Again, document all your attempts.

- From this point on, it's up to you. Start putting together a coherent picture of your conclusions, decide what is important and what is not so important, and assemble the important things. Also start thinking of the **best way to present your data**.
- Don't forget to SAVE OFTEN!