The Dirty Dozen on Doing Data Analysis with the Dynamic Duo

Good resources: <http://www.statsoft.com/textbook> <http://www.researchconnections.org>

1) Prepare for data analysis as you would a vacation or a business trip

2) Make good use of the online Helpguide within Excel/spreadsheet and SPSS (PSPP, STATA, SAS, R, Gretl, EPI-Info, or any other statistical software)

3) Make a codebook for the data and then start with entering the data using a spreadsheet file for further analysis (a codebook has four things for each variable: a)variable name, b)variable description, c)listing of all categories, and d)listing of all corresponding codes along with their units, dates, and the data source)

4) Specify what your data analysis goals are—consult your theory/hypotheses

5) Remember the basics for data entry purposes

a) Each column contains a different variable

b) Each row is a case (a unit of your analysis-person/family/agency/city/encounter...)

c) All variables should be set up so that the qualities of mutual exclusiveness,

exhaustiveness, high reliability, and high validity are upheld in the categories

d) Use numbers to stand for categories

e) It's better to have more categories to start with than too few--you can always combine

categories at a later time in the analysis

f) Code all dichotomous vars “0” for absence/false/disagree, “1” for presence/true/agree

6) Descriptive statistics like frequencies, percentages, rates, ratios, sums, means, median, and standard deviations can be done with Excel. You may want to use a spreadsheet to get an initial feel for the data before you use other packages.

7) Save the data as a spreadsheet, close the file, and then use other packages if desired

8) Within other packages, create "window dressing" (i.e. variable labels and value labels)

9) Recode, compute, or further refine cleaned data and save the file

10) Save whatever changes you've done in SPSS within a SPSS ".sav" file as well

11) Click on "Analyze" in order to answer data analysis questions you may have

a) Make conceptual/path models in order to represent what you are investigating

b) Use bivariate (crosstab) tables 1st in order to test ideas you have about the data

12) Be patient--the data analysis "ordeal" often requires several attempts and rethinking on your part--allow for several attempts in order to complete the analysis part of any project

13) Constantly empathize yourself as an outsider to your project. Break things down as simply as possible. Create graphs. Make explicit what you did do and what may have done better or differently.

***Final Points about Data Analysis:***

1) The use of stats or data analysis is a way to help make decisions that enable us to:

* choose a problem "worthy" of attention,
* find an answer/solution that is appropriate/effective,
* determine if progress has been made in addressing some concern, and
* share information or communicate ideas.

2) Stats in of themselves don't lie, only the folks who misrepresent them do.

3) Data analysis or stats as a tool is one part in the bigger picture of scientific things. Indeed, it should neither be forgotten nor thought of as the only answer.

4) The entire research context of the stats is just as important as the #s themselves.