Documenting Your Statistical Printouts

In the era of "clicky box" user-friendly statistical programs, it is often difficult to remember the origin of a particular printout -- what data set, what subsample of observations within the data set, what transformations of variables, etc. Here are some recommendations for keeping track of what you did. Most of this advice is just *fundamental scientific practice.* Along with the steps described here, you should keep a permanently bound lab notebook with notes in ink of what you did and when you did it. Recording research protocols is not covered here -- my job is to teach you to keep your data analyses straight.

1. Type the date into the output file (including the year) so it is printed right on the results. If you forget to type it in before printing it, write the date by hand when it is still hot out of the laser printer.
   -- Why include the year? Believe it or not, you might be working on publishing these results for longer than a year.
   -- Why date it at all? (1) Dating results is fundamental scientific practice. (2) You may find an error in your data file, correct it, and rerun the same analysis. Which printout is the correct one?

2. Keep a record of all data transformations. It is not enough to say to oneself, "I know this variable is the sum of two others because I named it 'XPLUSY'." If you transform data with a program you wrote (C, Pascal, etc.), save the code and the date you ran it, and document the output file. If you transform in a statistical package like SPSS, "Paste" your syntax, run the transformation from the syntax file, and save the syntax file. The "Paste syntax" box is the most important clicky-box in SPSS! I like to print copies of those syntax files and save them in my notebook for the project. You should also copy the appropriate syntax file into the output file so that it is printed with the output. All of these steps will allow you to re-trace your steps and find errors (or verify that you have done what you wanted to do correctly). Remember, a typo (or erroneous click) in the name of a variable when computing a new variable implies that the new variable is not what you think it is.
   -- In SPSS, I almost never use the "recode" option because the program will alter your original data without warning you. My recommendation is that you always create a new variable in SPSS when transforming your data. This advice is sound for other programs as well.
   -- Some programs will create temporary variables from your transformation instructions and put those instructions in the output file. I like this best because what you did is written right there in the output file. You can look at it whenever the printout is in your hand.

3. What data file was used should be stated as part of the output file. (SPSS does not automatically do this.) Type this in.
   -- Did you filter the data to omit certain cases? Perhaps you eliminated outliers, perhaps in separate runs you used different criteria for eliminating outliers. Document this on the printout while you are working, not later. One way is to create a separate data file for each different filtering and document what is in each data file. This makes it possible to re-trace one's steps fully to verify an analysis.

4. Page numbers, who needs page numbers? How about stapling or a three-ring binder? Some programs (SPSS) do not automatically put numbers on the pages when the output file is printed. The problem is that when you run several analyses that are similar, you can mix up the pages ("Is this the post-hoc test for analysis A or analysis B?"). In my lab we have two methods of keeping analyses together: (a) staple the pages hot out of the laser printer, (b) three-hole punch and put the analysis in a ring binder, c) electronic filing with comments embedded.. In both cases the analysis should be labeled as described above.

5. Electronic filing tips: Some people don't print statistical output files, but just file them electronically. If you do that, the documentation above is even more important. Plus, be careful if you open two such documents in your computer simultaneously. Flipping through pages on a screen can be confusing -- are you sure that the Fs and other statistics you are copying into your manuscript are the correct ones? Embedded comments will help you greatly.