

Introduction to SAS, mostly on PC's

Philip Dixon

25 July 2007

SAS Principles

SAS follows a 'write a program / run a program' model. There are a point-and-click programs called SAS/INSIGHT and SAS/ANALYST for simple analyses, but they will not do everything we need to do in this class. To get data analyzed using SAS, you must:

1. Write commands that tell SAS how to read the data
2. Write commands that tell SAS what sort of analyses you want
3. Tell SAS to execute those commands
4. Look at the .log file or LOG window to see whether SAS understood your commands
5. If you made a mistake, you need to:
 - (a) edit your program
 - (b) resubmit all or parts of it
6. Look at the output and interpret it.

You can use SAS in two very different ways.

- Windowing versions of SAS: (many on-campus PC's)
Most of you will use windowing SAS (PC's, some MAC's). After you start it, SAS will provide a screen full of windows. Your commands go in the program editor window. The log and output appear in the log and output windows.
SAS is a huge and expensive program. The license is an annual rental, not a purchase. Departments or individuals can get a license to install it on an on-campus computer. The Agronomy Department has licenses for lease, for about \$100 per year.
General information about SAS on PC's, Mac's and Linux is available at <http://www.stat.iastate.edu/>
- SAS on sas.iastate.edu
ISU maintains a dedicated SAS server (sas.iastate.edu) for campus use. This can be accessed in batch from off campus using putty or ssh (secure terminal programs, available through Scout). If you use 'batch' SAS, your commands go in a file with a .sas extension. The log and

output appear in files with .log and .lst extensions. This requires learning some UNIX. The SAS server is especially useful for long or slow programs.

You can also use the sas server in windowing mode at home if you have a high-speed internet connection. You need to install VPN, Xwin32 and/or Putty. All are available through Scout, the ISU online distribution manager of site-licensed software. Detailed instructions are at <http://www.stat.iastate.edu/resources/software/sas/linux-pc.html>

SAS programs

SAS programs are organized into:

- DATA steps. These convert your data file(s) into a form that is usable by SAS.
- PROC steps. These ask SAS to run a particular analysis.

You may have more than one DATA step and more than one PROC step. The typical 401/402/500 program has one DATA step and multiple PROC steps.

SAS is a very powerful database manager. We will use only the simplest parts of its data management capability.

Data structure: In this course, we will assume that a data file has 'flat-file by rows' format, i. e.:

- Each row contains one observation.
- Data values are separated by spaces (one or more).
- There may be a header line giving the variable names. This can be omitted.

Here is an example:

```
group yield
a 29.9
a 11.4
b 26.6
b 23.7
a 25.3
b 28.5
b 14.2
b 17.9
a 16.5
a 21.1
b 24.3
```

There are 11 observations (one per row). Each observation has 2 variables. The first identifies a treatment group; the second is the yield.

Data sets: SAS does not work directly on the original data file. The analysis parts of SAS (the proc steps) use data that are stored in a SAS data set. So, the first part of a typical SAS program creates a SAS data set from the original data. The minimum DATA step has a DATA line, an INFILE (or a CARDS line or a DATALINES line) and an INPUT line.

```
data tomato;
  infile 'C:\Documents and Settings\pdixon.IASTATE\My Documents\tomato.txt';
  input group $ yield;
run;
```

This creates a SAS data set called tomato. The data come from the file 'tomato.txt' in my My Documents folder. SAS works on many different operating systems, so it is not optimized for Windows. This means that you need to specify the full path to a file. That's the C:... stuff. The details of this change with different versions of windows. The easiest way to get the details correct is to highlight a file name, right click on the mouse, and select properties. You want the information labelled 'Location:'. I highlight it, copy it to the clipboard, then paste it into my SAS program. Remember, you also need the file name (tomato.txt).

Each row has two values. The first value will be stored with the variable name GROUP; the second will be called YIELD.

By default, SAS treats all variables as numeric. That is fine for YIELD, but the values of GROUP are not numbers. You tell SAS to read a variable as a character variable using the \$ **after** the variable name on the input line. If you omit the \$, you get lots of errors (value not numeric) in the log window.

PROC steps: Each proc step tells SAS to perform a particular analysis. Each proc step includes commands that describe the specific analysis. I will introduce appropriate commands as needed throughout the semester.

```
proc ttest;
  class group;
  var yield;
  title 'T-test of tomato yield';
run;
```

```
proc boxplot;
  plot yield*group;
  title 'High resolution boxplots for each group';
run;
```

Each command to SAS MUST end with a ; One command can span multiple lines, or one line can have multiple commands. SAS only cares that each command ends with a ;

HINT: If SAS gives you a bunch of errors, and the commands look correct, check to see if you left out the ; This is a common problem, and omitting the ; really confuses SAS.

While not necessary, it is good practice to end each PROC or DATA step with a run; command.

HINT: If you submit some SAS commands but don't get any output. Or, you don't get any output from the last proc; step. You forgot the run; command. You don't need to resubmit everything. Just type run; in your program editor window (probably at the end), highlight it, then click submit. You should get the output from your last proc; step.

There is almost always more than one way to do something in SAS. I've suggested my approach. If you see a different PROC used or a different sort of DATA step, it may just be a different way to accomplish the same thing.

More details on SAS data step programming and basic SAS proc's are in the second handout.

Working with windowing SAS on a PC

To start SAS: Click on the SAS icon.

The windows:

Four windows will appear on your screen.

- Explorer: Provides a point-and-click interface to SAS data sets. This is less important and I usually ignore it.
- Editor: You enter your SAS commands here. When you give the SUBMIT command, SAS will execute the commands that are in this window. You can execute a subset of the commands here by highlighting the desired commands before using SUBMIT.
- LOG: This window contains SAS's responses to your commands. Any errors will appear here. Various informative messages will appear here.
- OUTPUT: This window contains the output from your SAS commands.

editing files on PC's, MAC's The SAS editor window works just like a word processor. Type your commands in and use the enter key to get a new line.

If you use Word or other standard word processor to edit your file(s), you need to save the files in text format (.txt) using SaveAs. SAS can not read files saved as Word documents (.doc files).

Loading a class program The class web site will have examples of SAS programs. You can download them using Internet Explorer, save them, then load them into SAS using FILE/OPEN.

The default extension for a sas program is .sas. Your browser may not allow you to save the file as a .sas file. You may have to save the file as a text file (.txt). Then, rename the file (highlight the file name, right click, rename) and replace the '_sas.txt' with '.sas'. This is a windows/Internet Explorer problem.

Saving, Runing, Exiting To save the contents of any window to a file, under the *file* menu in the desired window, click on *save as...*, type in (or click on) where you want to save it, name the file, and click *OK*.

Reminder: always check the pathname to make sure it will save it where you want to save it. The default directory is the SAS program directory, which may not be easy to find.