

Statistical Methods and Computing, 22S:30/105

Instructor: Cowles

Lab 1

Jan. 28, 2009

1 Using SAS outside of ITCs

You can access SAS from off campus by using the ITC Virtual Desktop. Go to

itc.uiowa.edu

and click on Virtual Desktop. You will be prompted to assent to an agreement and to enter your Hawkid and password. Select the SAS 9 folder, and from it, SAS 9.2. SAS will come up on your screen. When you are finished, you must both exit from SAS and log off from the Virtual Desktop. When you run SAS this way, although you can interact with it on your screen, it is actually running on a computer on campus. Consequently, SAS can't access files that are on your hard drive unless you specify drive letter mapping. In this lab, I will show you the most reliable way to get data into SAS when you are running it on the Virtual Desktop.

If you want to actually install SAS on your own computer, SAS for Students, Windows version, is available at a reduced cost of \$28.50 for students using the software for academic purposes. This license needs to be renewed in October. Available at the IMU Bookstore. See the SAS Information page,

<http://cs.its.uiowa.edu/software/sasstudenttudent.shtml>

for product information.

2 Getting started in the ITC

ITC computers will display a login screen. Users should enter their HawkID and password in the spots provided. Students can find their HawkID and default passwords on the ISIS system.

3 Downloading files from the course web page

Bring up a web browser, either Firefox or Internet Explorer Enter the address of my web page in the location box.

www.stat.uiowa.edu/~kcowles

Then click on "Course homepages" and "22S:30/105."

Click on "Datasets," and when the next screen appears, click on the underlined link "Datasets." Three types of files may be accessed:

- Files ending in .dat are data files for your use with the software package SAS.
- Files ending in .info contain descriptive information about datasets.
- Files ending in .txt are datasets for use in a different class with a different software package.

Note that this list is case-sensitive, with all files with names beginning with capital letters appearing before all other files.

If you are going to work with SAS on the Virtual Desktop, it is easiest to copy and paste the data from the web page directly into the SAS program. If you are running SAS on the computer at which you are sitting, it is preferable to download the dataset into a file and tell SAS where to find the file. Below, I will show you how to have SAS access data both ways.

To download a data file for use in this lab:

- Right click the file name (use the *right*, not left, mouse button)
- In the dialog box that opens, left click "Save target as" (use the *left* mouse button)
- In the "save as" dialog box chose My ComputerLocal Disk C:/ Temp. Click "Save."
- (If you preferred to save the file on your own disk in drive A: so that you could use it on a different computer later, you would move to drive A: in the dialog box before saving.)

Left click the "billion.info" file to read a description of the "billion" dataset. Click "Back" to get back to the list of datasets. Then download the file "billion.dat" according to the directions above.

4 Other useful features on the course web page

Return to the main course web page and click "Web resources." Note that there is a directory of the locations and hours of all the campus ITCs, as well as a link to the Mathematical Sciences Library electronic reserve (where solutions will be posted).

Lecture notes, homework assignments, and lab handouts are posted under "Handouts." These are in a format that may be read and printed in most ITCs.

5 Accessing SAS

Click on Start/All Programs/SAS/SAS 9.1 (English). Alternatively, you may access SAS on the Virtual Desktop as described above.

In either case, you will get a screen that shows:

- a menu bar
- a log window
- a program editor window

6 Entering commands and programs

Click in the program editor window. You may now type commands and programs in this window.

7 How SAS programs and commands are organized

Use a *DATA step* to organize your data by creating a SAS dataset. Then use *PROC steps* or automated features to analyze your data. Once you have created a SAS dataset, you may apply any SAS procedures or automated features to it during the SAS session without recreating the dataset.

DATA and PROC steps consist of SAS *statements*. Each statement must end with a semicolon. Most statements include one or more *keywords* that must be spelled exactly as shown.

8 The DATA step: Creating a SAS dataset

Before it can process data, SAS must read in the data in the form of a table with

- a row for each *observation*
- a column for each *variable*

You must choose a name for the entire dataset and a name for each variable. SAS has the following rules for names:

- SAS names must begin with a letter or an underscore. The remaining characters in a SAS name can be letters, numbers, or underscores. There must be no embedded blanks.

SAS distinguishes between two types of variables: *numeric variables*, which contain only digits and decimal points and with which arithmetic operations may be done, and *character variables* (all other kinds of data).

9 Controlling print width

Put this line at the beginning of every SAS program if you want output to print correctly on 8-1/2 by 11 inch paper:

```
options linesize = 75 ;
```

10 Reading data in from an existing datafile

You have saved the file “billion.dat” in the “temp” directory. Use an “infile” statement to tell SAS to use it.

```
data billion ;                * gives dataset a name for SAS ;
infile 'c:\temp\billion.dat' ; * tells SAS where the data is ;
input wltch age region $ ;    * names the variables in each row ;
                              * $ after region identifies character vbl ;
run ;                          * end of data step ;
```

11 Copying data directly into the data step

In this alternative method to get data into SAS, do not use an “infile” statement. Instead use the “datalines” statement, and copy and paste the data in immediately following it. Put a semicolon all by itself on the line after the last line of data.

```
data billion ;                * gives dataset a name for SAS ;
input wltch age region $ ;    * names the variables in each row ;
                              * $ after region identifies character vbl ;

datalines ;
37 50 M
24 88 U
14 64 A
.
.
.
1 59 E
1 . E
1 . 0
;
run ;                          * end of data step ;
```

Type one of these data steps into the program editor window. To make SAS run these statements and create the dataset, use the mouse to highlight the block of statements and then click on the icon of the running man.

SAS will use the log window to tell you what it has done. Be sure to check the log window for any error messages. If any errors are reported, click in the program editor window to make it active. Correct the errors in the code and then rerun the block of code.

Note: if you wanted to read in the file from your own disk in the A: drive the infile statement would be:

```
infile 'a:billion.dat' ;
```

12 Using SAS procedures to list and tabulate the dataset

Once the dataset is created, you may run SAS *procedures* to analyze it. To list the entire dataset:

```
proc print data = billion ;
run ;
```

To get a frequency distribution of the regions in which billionaires lived:

```
proc freq data = billion ;
tables region ;
* tables is a keyword; region is the name of
* the variable for which you want counts ;
run ;
```

The output is:

The FREQ Procedure

region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	38	16.31	38	16.31
E	80	34.33	118	50.64
M	22	9.44	140	60.09
O	29	12.45	169	72.53
U	64	27.47	233	100.00

13 Proc univariate: SAS workhorse of descriptive statistics

Use *proc univariate* for quantitative variables when you want the following:

- means
- medians
- quartiles
- 5-number summary

- stem plots (for small datasets) or histograms (large datasets)
- boxplots

```
proc univariate plot data = billion ;
var wlth ;
run ;
```

The output is:

The UNIVARIATE Procedure Variable: wlth

Moments

	N	233	Sum Weights	233
Mean	2.68154506		Sum Observations	624.8
Std Deviation	3.31884032		Variance	11.0147011
Skewness	6.57544276		Kurtosis	56.9655987
Uncorrected SS	4230.84		Corrected SS	2555.41064
Coeff Variation	123.765972		Std Error Mean	0.21742446

Basic Statistical Measures

	Location		Variability
Mean	2.681545	Std Deviation	3.31884
Median	1.800000	Variance	11.01470
Mode	1.000000	Range	36.00000
		Interquartile Range	1.70000

Tests for Location: Mu0=0

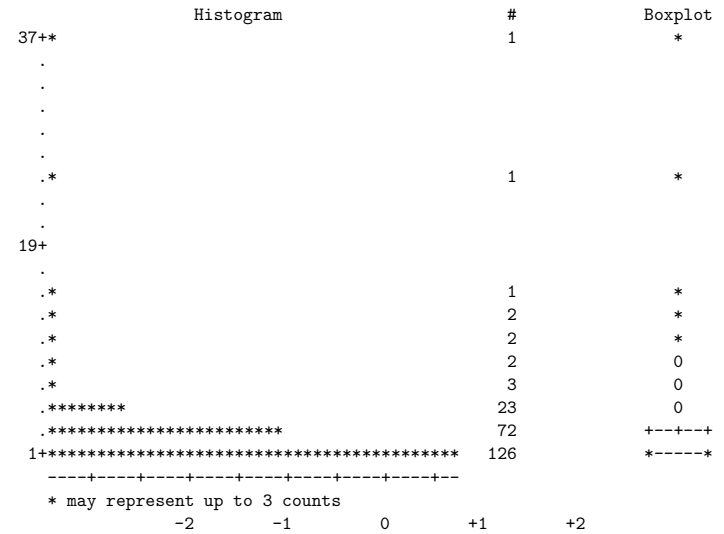
Test	-Statistic-		-----p Value-----
Student's t	t	12.33323	Pr > t <.0001
Sign	M	116.5	Pr >= M <.0001
Signed Rank	S	13630.5	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	37.0
99%	14.0
95%	6.2
90%	4.5
75% Q3	3.0
50% Median	1.8
25% Q1	1.3
10%	1.1
5%	1.0
1%	1.0
0% Min	1.0

Extreme Observations

---Lowest---		---Highest---	
Value	Obs	Value	Obs
1	233	13	4
1	232	13	5
1	231	14	3
1	230	24	2
1	229	37	1



14 Bar graphs and pie charts

```

goptions device = win ;
pattern v = solid color = gray ;

proc gchart data = billion ;
vbar region ;
title 'Billionaires in 1992; Regions ' ;
run ;

proc gchart data = billion ;
pie region ;
title 'Billionaires in 1992; Regions ' ;
run ;

```

15 SAS Analyst: Automating descriptive statistics and plots

Use the following steps to get into “Analyst” from the menu:

- Solutions
 - Analysis
 - * Analyst

You must specify which dataset you wish to use. Do so by clicking

- File
 - Open by SAS name
 - * Work library (double click)
 - Billion (double click)

To get a bar chart of the counts of how many billionaires are from each region, select “Graphs/Barchart/Vertical.” Then click on “REGION” and then on “Chart” to indicate that the region variable is the one you want graphed. Click “OK.” Then click on “Options,” check off “Bars for each discrete level,” and click “OK.” Then click “OK” again.

To get a numeric and graphical summary of the values of the wealth variable, select “Statistics/Descriptive/Distributions.” Click “WLTH,” then “Analysis / plots.” Check off the types of plots you want.

16 Printing and Saving Files

Copying output from SAS windows into Microsoft Word will enable you to edit the SAS output and incorporate it into your homework writeups. You can then print from Word. When you highlight a block of text in the SAS output window in order to copy it, do not highlight all the way to the right margin of the last line. Due to a bug in SAS, that prevents the copy from working.

To save a file, click in the window whose contents you want to save. Go to the file menu and choose “Save as”. SAS will automatically give the file extension “.sas” to SAS commands and programs. For example, to name a SAS program “myprog,” you would type

a:myprog

in the box for the name of the file.

If you’re running SAS on the Virtual Desktop and want to save a program for future use, an option is to copy and paste your SAS code into a text document on the computer on which you’re working, and save the text document. Then at your next SAS session, you can bring up the text document and copy and paste the code into the SAS program editor window.

17 When you have finished...

Be sure to exit from SAS using the File menu, and to log out of the computer using the icon on the desktop.