SAS for Data Management, Analysis, and Reporting
225:172, Summer 2003
Instructor: Coates
Lab 1
Jun. 15, 2004

1 If you want SAS for your own computer

SAS for Students, Windows version, is available at the IMU Bookstore at a reduced cost for students using the software for academic purposes. This license needs to be renewed in October. See the Software Listing page:
http://www.its.uiowa.edu/cs/software/softwarelistings.html
for pricing information. See the SAS Information page,
http://www.its.uiowa.edu/cs/software/sas.html
for product information.

2 Getting started in the ITC

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

3 Downloading files from the course web page

Click on Start/All Programs/Internet Explorer. Enter the address of my web page in the location box at the top of the Explorer screen:

www.stat.uiowa.edu/~jcoates

Then click on “Course homepage” and “225:172.”

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 use with the software package SAS.
• Files ending in, info or, chik contain descriptive information about datasets.
• Files ending in, dat are datasets for use in a different class with a different software package.

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 choose My Computer/Local 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.

To read a description of the Davis dataset, left click the “Davis,chik” file. Choose to open the file from its current location. If you receive the message that Windows does not know how to read this type of file, choose to “Select from list” and then choose WordPad from the list of software. Click “Back” to get back to the list of datasets. Then download the file “Davis.dat” according to the directions above.

Similarly, read the description in the “poverty.info” file, and then download “poverty.dat.” Note that the computer on which these datasets are tied alphabetizes with all capital letters first, followed by all lower case letters.

4 Accessing SAS

Click on Start/All Programs/The SAS System/SAS 9.0.

You will get a screen that shows:

• a menu bar
• a log window
• a program editor window

5 Entering commands and programs

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

6 How SAS programs and commands are organized

Use a DATA step to organize your data by creating a SAS dataset. Then use PROC steps to analyze your data using SAS procedures. Once you have created a SAS dataset, you may apply any SAS procedures to it 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.

7 The DATA step: Creating a SAS dataset

Before it can process data, SAS must read 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 arith-
metical operations may be done; and
- *character variables* (all other kinds of data),

8 Controlling print width and character formatting

Put these lines at the beginning of every SAS program if you want output to print correctly
on 8 1/2 by 11 inch paper and to base lines print correctly in tables and graphs:

```
options linesize = 75;
formchar = '1 1=' = /<c**';
```

The character string for formchar is included on the course web page under “Datasets” in
the file called “formchar.” You may copy it from there into your program.

The linesize option tells SAS how many characters to print on each line of text. For
normal text, a maximum of 80 characters can be printed per line. The formchar option
tells SAS what characters to use to print the lines dividing cells in certain kinds of tables.
If we let SAS use its default setting for formchar, these tables will not print correctly.

9 Reading data in from an existing dataset

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

```
data Davis;  * given dataset a name for SAS
  infile 'c:\temp\Davis.dat';  * tells SAS where the data is
```

input subj sex $ mgst shgt rvgt rhgt;  * name the variables in each row
  $ after sex identifies character vbl
  * end of data step

Type these lines 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:\Davis.dat';
```

10 Using SAS procedures to list, explore, and check the data

Once the dataset is created, you may apply SAS procedures to it. The first step is to make
sure the dataset was read in correctly. To list the entire dataset:

```
proc print data = Davis;
run;
```

To print only the first few lines of the dataset:

```
proc print data = Davis(obs = 12);  * number of observations to list;
run;
```

The first step check for validity of binary, nominal, and ordinal variables is to run a fre-
quency table to check for missing values or invalid values. To get a frequency distribution of
the gender of the subjects:

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

11 Using proc means for simple data description and data checking

A simple procedure to get a feel for the sample distributions of quantitative variables is
proc means.

```
proc means data = Davis;  * default output;
var mgst shgt rvgt rhgt;  * list of quantitative variables to summarize;
run;
```
proc means data = Davis n mean std ev min median max ;
  * options to specify descriptive statistics ;
var weight height weight ;
run ;

You must use a more detailed procedure to discover which observations contain outlying values. Proc univariate is the workhorse of descriptive statistics for quantitative variables. Use proc univariate for quantitative variables when you want the following:

- mean
- medians
- quartiles
- 5 number summary
- stem plots (for small datasets) or histograms (large datasets)
- boxplots
- to identify outlying values

proc univariate plot data = Davis ;
var weight height weight ;
run ;

The primary purpose of Davis's study was to determine whether the self-reported variables weight and height could be used as substitutes for the more difficult to obtain measured values weight and height. So before carrying out analyses, we should check whether there are any observations in the dataset that have unreasonable relationships between the pairs of variables. Let's make a rough scatterplot. Which variable should go on the Y axis, and why?

proc plot data = Davis ;
plot weight * height ;
  * plot response=var * explanatory=var ;
run ;

This type of graph is easy to print and can be printed on any kind of printer. It is good for homework and planning of reports and projects. We will learn to make much more professional looking plots later for use in presentations and reports.

What might be going on here?

Try a similar plot for the two weight variables.

12 Working with a more complex dataset

The following data step will read in the "poverty" data. Recall that the country name is character variable. Different names have different numbers of letters. We need to use a length statement to tell SAS how many letters are in the longest name.

data poverty ;
input 'c:\temp\poverty.dat' ;
length country $ 20 ;
skip linebirth dthrate ifdth lifeexpm lifeexpf popgap group country ;
run ;

Next we will use the id statement in proc print to make the country name appear as the first column in the listing and to suppress the observation numbers.

proc print data = poverty (obs=25) ;
id country ;
run ;

We can request specific descriptive statistics in proc means.

proc means data = poverty n mean std median max = 1 ;
run ;

We can use proc freq to check whether counts of the countries in each country group are reasonable. Note the use of the title statement in this procedure. The same title will be printed on the output of each procedure after this one unless we change it.

proc freq data = poverty ;
tables group ;
title 'Count of countries in each country group' ;
run ;

We can put options on the tables statement after a slash. The nocum option suppresses the columns of cumulative counts and cumulative percents.

proc freq data = poverty order = freq;
tables group / nocum ;
title 'Count of countries in each country group' ;
run ;

13 Working with a computed variable

Suppose that we were interested in the difference between life expectancies for females and life expectancies for males in the countries. We can add a statement to the data step to compute this difference for each observation in the dataset. Add the line, and then run the entire data step.

Try a similar plot for the two weight variables,
data poverty;
infile c:\temp\poverty.dat;
length country 8 20;
input livexp @ 10;
input lifeexp @ lifeexpf pgrp@ group country;
expdiff = lifeexpf - lifeexp;
run;

Now we can treat the new variable in the same way as any other quantitative variable in the dataset,

proc means data = poverty;
var expdiff;
run;

14 More on proc means and proc univariate

We can use the id statement in proc univariate to make SAS print the values of an identifying variable alongside the extreme values of the quantitative variable being summarised. Let's see which countries have the extremely low and extremely high values of male life expectancy. Do these seem plausible, or do you suspect some data errors?

proc univariate data = poverty;
var lifeexp;
id country;
run;

Here is an example of producing proc means output separately for the observations in groups defined by the value of another variable.

proc means data = poverty;
var group;
var lifeexp;
title 'life expectancy for males by country group';
run;

15 Bar charts

proc chart data = poverty;
var group/discrete;
title 'Bar graph of number of countries per group';
run;

16 Printing and Saving Files

Copying output from SAS windows into Microsoft Word, and then printing from Word is your best bet for printing sections of SAS code and output. SAS has a silly bug in order to copy text from a SAS window, highlight the section you want, but don't highlight all the way to end of the last line.

To save a file from SAS to disk, click in the window whose contents you want to save, Go to the file menu and choose "Save as." If you want to save your SAS program and work with it again later, you must save it on your disk in the C drive, SAS will automatically give the file extension "sas" to SAS commands and programs. For example, to name a SAS program "myprog," you would type

myprog

in the box for the name of the file.

If you just want to temporarily save a file and not need to use it again after you leave the lab, you may save it to the "temp" directory.

See page 19 of the textbook for directions on printing or saving just part of the program output directly from SAS.

17 Using SAS online help

Click on "Help" in the pull-down menu, then on "SAS Help and Documentation" then on "Help on SAS Products," Documentation of the data step and of the procedures that we have used for simple descriptive statistics and plots can be found under "Base SAS." Click on "Base SAS Procedures" to see a list of all the procedures included in Base SAS. Then read the documentation for proc means.

Later on we will use procedures in SAS/Stat and SAS/Graph.

18 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.