**Statistical Methods and Computing, STAT:2010/4200**
Lab 1

# 1    A collection of datasets for this course

Bring up a web browser, and Enter the following address in the location box.

`http://homepage.stat.uiowa.edu/~kcowles/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. Simply left click the file name.

- (optional) 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. To download a data file:

  - 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, say, 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.

## 2    Accessing SAS

If available, click on Start/All Programs/SAS/SAS 9-4.

Alternatively, you can access SAS from on or off campus by using the ITC Virtual Desktop.
Go to

```
http://virtualdesktop.uiowa.edu
```

Log in with your Hawkid and password. Single click SAS 9-4 and wait patiently. SAS will
come up on your screen.
(If you are using virtual desktop on your computer for the first time, you might encounter
a screen that asks you to download and install the latest version of Citrix. Then you might
also be asked to setup an account using your uiowa hawkmail, before you can launch SAS.)

You will get a screen that shows:

- a menu bar

- a log window

- a program editor window

## 3    Entering commands and programs

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

## 4    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 semi-
colon. Most statements include one or more *keywords* that must be spelled exactly as
shown.

## 5    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).

# 6   Controlling output format

In SAS 9-4, the default output is in html format. This can be changed in the menu bar on top of the window, by clicking Tools/Options/Preference, under the Results menu,

- When the box "Create listing" is checked, text output will be produced that are easier to copy and paste.

- When the box "HTML" is checked, better looking graphs and tables are produced.

You can check one or both of the boxes, depending on your need.

# 7   (Optional) Reading data in from an existing datafile

We mention how to read data from an existing datafile in this section.

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 wlth age region $ ;            * names the variables in each row ;
* $ after region identifies character variable ;
run ;                                * end of data step ;
```

If you do this in virtual desktop, you need to find the file you saved in the Explorer located on the left side of SAS window. For instance, you saved "billion.dat" in the desktop of local C drive, you may click This PC - Local Disc (C) - Users - ShengWang - Desktop. Then right click billion.dat to see its properties, where we could find the path. You will likely find a path that is similar to the following one and use it in you infile statement instead.

```
'\\Client\C\$\Users\ShengWang\Desktop\billion.dat'
```

# 8   Copying data directly into the data step — always applicable, but inconvenient if the dataset is very big)

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 wlth age region $ ;              * names the variables in each row ;
                                       * $ after region identifies character variable;
datalines ;
37 50 M
24 88 U
14 64 A
.
.
1 59 E
1 . E
1 . O
;
run ;                                  * end of data step;
```

Type the above 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.

# 9  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 |

## 10    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 ;
```

## 11    Bar graphs and pie charts

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

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

## 12    Printing and Saving Files

Copying output from SAS windows into Microsoft Word (or other text editors) will enable you to edit the SAS output and incorporate it into your homework writeups. You can then print from Word. Recall from section 6 that for tables, text output may be easier to copy and paste.

To save a 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.

- If SAS is physically installed on the computer, you can (1) clicking in the window whose contents you want to save (2) go to the file menu and choose "Save as". SAS will automatically give the file extension ".sas" to SAS commands and programs.

## 13    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.