1 Data set to download

Please download the datasets OECD.dat and billion.dat from the course web page after reading their associated .txt files.

Read the OECD dataset into SAS,

```sas
data OECD;
  infile 'c:\temp\OECD.dat';
  input country $13, pgdp pch beds los dose ismfat ;
run ;
```

2 Analyst for regression

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)
      - OECD (double click)

To create a scatterplot, choose "Graphs/Scatterplot." Use the interactive window to specify the explanatory variable on the X axis and the response variable on the Y axis.

To do regression analysis, choose "Statistics/Regression/Simple." Again, interactively specify the explanatory and response variables. Other choices in the window can be used to request predicted values and specific plots. If you request a scatterplot of observed vs. independent, the resulting plot will have the regression line drawn on it. To copy a graph from Analyst into a Word document, click on the graph, then use the pull-down Edit menu to select "Copy graph." You can then "Edit/Paste" into a Word document.

To get out of Analyst and back to regular SAS, use the pull-down File menu and choose "Close."

3 Insight for regression

Insight is another point and click facility built into SAS. We will be using its graphical features later on when we study multiple regression. In case you want to try it now, here are some instructions.

From the main pull down window, select the following sequence of choices:

- Solutions
  - Analysis
    - Interactive data analysis

In the window that appears, you must specify which dataset you wish to use. Do so by clicking

- Library: Work
  - Dataset: OECD
  - Open

To do regression in Insight, choose

- Analyze
  - Fit

To identify the response variable, use your mouse to click "PCGDP" and then "Y." Similarly, copy "PCGDP" into the "X" column. Click "OK" and lots of regression output and plots will appear.

To get out of Insight and back into command mode, click in the window showing the data in spreadsheet form. Then pull down the "File" menu and choose "Exit."

4 Using formats to get SAS to print something other than the values a variable actually contains

We will be using the billionaire dataset again today. Its variable called `region` contains abbreviations ("A" for Asia, "E" for Europe, etc.). If we want SAS to print out the complete words instead of the abbreviations, so that tables and graphs are more understandable, we need to run a "proc format" before the data step. The data step must then refer to the formats defined in the format procedure.

```sas
options linesize = 76 ;
proc format ;
```
value $regfmt 'A' = 'Asia'  
'E' = 'Europe'  
'M' = 'Middle East'  
'O' = 'Other'  
'U' = 'US' ;
run;

Because the original data values in the region variable are characters rather than numbers, we have to use a dollar sign as the first character in the name of the format.

Note the format statement in the data step below. It tells SAS to apply the format you have defined here to a particular variable. When you use the format statement in a data step you must put a period at the end of the format name.

5 Using labels to get SAS to print more descriptive variable names

data billion ;
 infile 'c:\temp\billion.dat' ;
  input wealth age region $;
  format region $regfmt ;
  label wealth = 'Wealth in Billion $' ;
  age = 'Age in Years' ;
run;

Now enter and run the following code to see how the formats and labels affect the output of the "print" and "freq" procedures.

proc print data = billion (obs = 20) ;
run ;

proc print label data = billion (obs = 20) ;
run ;

proc freq data = billion ;
tables region ;
run ;

6 Formats for numeric variables

Formats can also be used to group numeric data. Add a list to your format procedure and change one line in the data step as follows.

proc format ;
  value $regfmt 'a' = 'low'  
'A' = 'low'  
'B' = 'medium'  
'C' = 'high' ;
run ;