

## Inputting data into SAS

- I. We'll first input some information for a sample of technology firms. The variables inputted are, in order, revenues (in millions), revenue growth, return on equity, total shareholder return and profits (in millions). The data is located at [techa.prn](#). The program to read the data is [prog1.sas](#).

### Notes:

- a. Always create a *data* statement before performing such tasks as reading in or manipulating data. You can do a number of things within an individual data step but you probably would want to separate your tasks into different data steps.
  - b. The *infile* statement directs SAS to look for an external data file at the specified location.
  - c. The input statement specifies the names of the variables inputted. The number of variables should normally equal the number of columns of the data inputted.
  - d. The *proc print* statement prints out the data on screen that you have read. Notice that there are a few missing values.
- II. Create a SAS program to read in the data set [techb.prn](#). The data set has two variables: revenues and profits. Use the *proc print* statement to look at what you have read in.
  - III. Create a SAS program to read in the data set [techc.prn](#). The data set has three variables: revenue growth, return on equity and total shareholder return. Use the *proc print* statement to look at what you have read in. Do you notice the missing values?
  - IV. Now we will take a data set that is not readable by SAS in its present form and convert it to a SAS-readable file. The data is the original Excel spreadsheet the above variables came from: [tech2.xls](#). Eliminate the non-numeric data in the file. We will create the program to read in the data.

### Notes:

- a. You have to convert the file into a text file that does not have tabs between the variables. Spaces should serve to separate the variables. Excel can create *formatted text - space delimited (.prn)* files, which can be read by SAS.
- b. Given that you are inputting the variable without a format statement, be sure that there is at least one space between every variable for every observation.
- c. Keep track of which variable is which! I rearranged the numeric data from what is used above.

- V. We next take a data set that is in html format on the web: [data](#)

This particular file was saved as html using Excel. So we can easily edit it with Excel. Delete the non-numeric variables except for the one representing company's country of origin. To input a non-numeric variable into SAS, add the \$ sign within the input statement after the non-numeric variable name.

- a. Create a variable indicating whether the company originates from the US or not.
- b. Calculate the mean and standard deviation in profits for US and non-US firms separately.
- c. Calculate the same statistics for profits but now distinguish between firms whose revenue growth was at least ten percent and those with growth below ten percent.

- VI. This [file](#) is already formatted to be read by SAS. The data represents characteristics of a sample of cities. The variables in order: city name, percent non-hispanic white, percent over age 65, percent of adults with college degree, percent employed in professional occupation, median household income.

- a. Input data with the city name variable included.
- b. Calculate the proportion of cities with median incomes over \$60,000.
- c. Calculate the mean percent employed in professional occupations separately for cities with median income over \$60,000 and less than (or equal to) \$60,000.

- VII. Download and set up the following excel data set to be read in SAS: [Library Data](#)

- a. Calculate the total number of children in each city.
- b. Perform the following regression (regression commands are found in [prog2.sas](#)):

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

y – total library expenditure (expend)

X<sub>1</sub> – number of residents (residents)

X<sub>2</sub> – dummy variable for city/county run library (citlib)

X<sub>3</sub> – percent school-aged children (school)

X<sub>4</sub> – median income (medinc)

- c. Run the log/log version of the regression model above.
- d. Run the log/linear version of the regression model above.

```
proc sort;by citlib;  
proc ttest;class citlib;var expend;
```

```
Data instruct1;  
infile "a:tech1.prn";  
input revenue revgrowth ereturn sreturn profit;  
proc print;  
run;
```

```
Data instruct3;  
infile " ";  
input cityname $ code expend residents citlib school medinc;  
Data instruct3a;  
set instruct3;  
proc reg;  
model expend=residents citlib school medinc;  
run;
```