

## 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 comma delimited data is [techa.csv](#).

### 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 *delimiter=','* statement indicates to SAS that the variables are separated by commas. SAS can input various data formats.
  - d. The input statement specifies the names of the variables inputted. The number of variables should normally equal the number of columns of the data.
  - e. The *proc print* statement prints out the data on screen. Notice there are a few missing values.
- II. Create a SAS program to read in the data set [techb.csv](#). 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.csv](#). 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 and make it more easily readable by SAS. The data is the original Excel spreadsheet the above variables came from: [tech2.xls](#). Eliminate the title on top and the variable names. Save as comma delimited. Keep track of what the variables are. To input a non-numeric variable into SAS, add the \$ sign within the input statement after the nonnumeric variable name.
  - V. We next take a data set that is in html format on the web: [data](#). This file was saved as html using Excel. We can easily edit it with Excel.
    - 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.
- Input data.
  - Calculate the proportion of cities with a median income over \$60,000.
  - 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](#). Perform the following regression.

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

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)