MET MA 603:
SAS Programming and Applications

Importing Data with the Data Step - Part II

Informat-Style Input

Recall the requirements SAS has for numeric data: only the digits 0 – 9, periods, +/-, and E are allowed.

When the data in an external file is formatted in some way ($ sign, commas, dates), SAS will not be able to import it as a number using Data Step methods (Proc Import is able to recognize a few types of formatting, but not all).

In these situations, the formatting used in the external file must be indicated, using what are called **Informats**.

Using Informats

Informats are included in the Input statement, after the name of the variable having the formatting.

```
data   city_pops9;
infile  "C:\Users\govonlu\Data\city_populations9.txt" ;
input   City $12. +1 State $2. Population comma9.;
run;
```

Character Informats begin with the dollar sign. They end with a period to help SAS distinguish them from variables.

With Formatted Input, the number of characters that SAS reads in for each variable are specified as part of the informat (similar to the fixed-width method). Unless otherwise specified, SAS starts reading in data from wherever it left off with the previous variable.

The notation +1, +2, etc. move the SAS cursor the specified number of place to the right.

Examples of Informats

Character

$*w*. $CHAR*w*. $UPCASE*w*.

Date and Time

DATE*w*.   MMDDYY*w*.

Numeric

w.   COMMA*w*.   DOLLAR*w*.

PERCENT*w*.   PERCENTN*w*.

Some numeric data allows for a decimal place. The number of characters to read in must account for the numbers after the decimal point, as well as the decimal point itself.

Combining Style

When using the Data Step to import, we are not restricted to only one style. Each variable can be read using a different style (List, Fixed-Width, Informat). In fact, this is often the most efficient way.

The most common errors that come with mixing input styles are in not understanding the location of the SAS cursor. Each style impacts the cursor differently.

Pointers are a useful tool for controlling the SAS cursor. Pointers have a similar function as the +n notation. The difference is that Pointers move the cursor to a specific column, while the +n notation move the cursor relative to its current position.

Practice

Import september_daylight.txt

Import aoi_curve.txt

Import scores6.txt

Import state_populations.txt

Readings

- Textbook sections 2.7, 2.8, 2.9, 2.10
- http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000204375.htm
 (note the explanation of the difference between the $CHAR and $*w* informats)