MET MA 603:
SAS Programming and Applications

Basics of Data Step Processing

Data Steps Processing

The underlying structure of a Data Step is a loop:

The keyword Data initiates the Data Step and creates an empty output dataset with the specified name.

Each statement is executed in sequence for each input observation.

The first time SAS encounters a new variable in the code, the variable is added to the dataset (the value is set to missing initially). Once a value is determined for a variable, the value is updated.
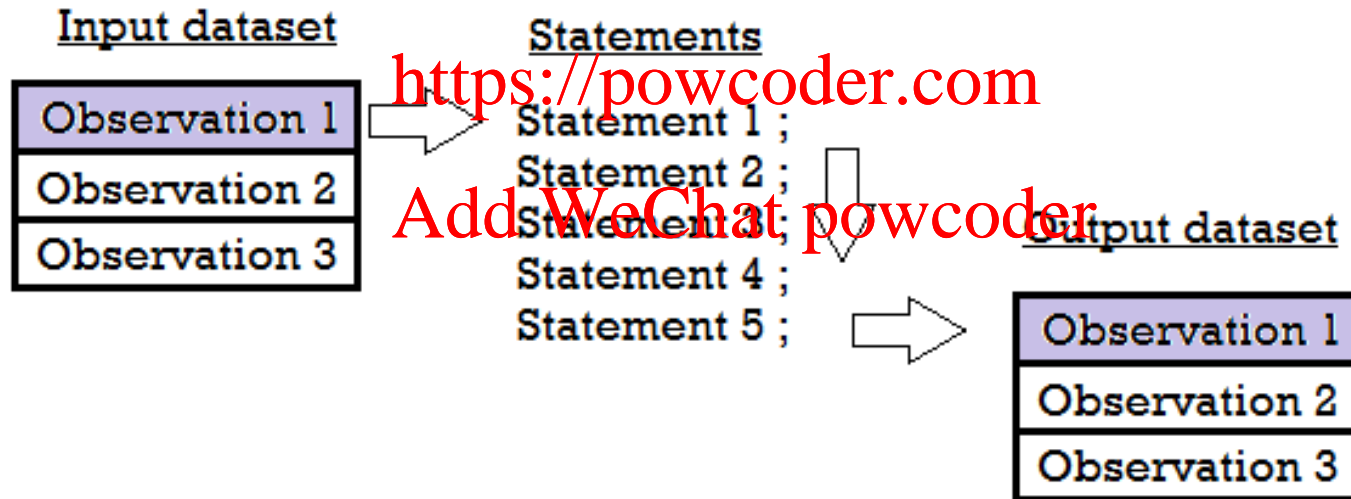
After executing the last statement, the result is written to the output dataset.

The loop is finished once SAS has processed the last observation.

Data Steps Processing (cont.)

Input dataset

Observation 1
Observation 2
Observation 3

Statements

Statement 1 ;
Statement 2 ;
Statement 3 ;
Statement 4 ;
Statement 5 ;

Output dataset

Observation 1
Observation 2
Observation 3

Set Statement

The **Set Statement** instructs SAS to process observations from an existing SAS dataset.

The two Data Steps below are similar. The first Data Step reads from an external file and the second Data Step reads from an existing SAS dataset.

```
data  city_pops1;
infile "C:\Users\govonlu\Data\
city_populations1.txt" ;
input  City $ State $ Population ;
run;

data  city_pops1_with_country;
set      city_pops1;
Country = "United States";
Population_000s = Population / 1000;
format Population_000s comma10.0 Population comma13.
;
run;
```

Dataset Options

Dataset Options allow for modifications to be made to the variables in a dataset.

Drop instructs SAS to remove the listed variables from the dataset.

Keep instructs SAS to remove all but the listed variables from the dataset.

Where instructs SAS to only keep observations that satisfy specified criteria.

Rename instructs SAS to change the name of a variable to whatever name is indicated.

Using Dataset Options

The Dataset options are entered inside parenthesis following the name of a dataset. They can be used with either the input dataset or the output dataset.

The dataset options can also be entered as statements. However, care must be taken to understand which dataset they operate on.

When entered as a statement, "Drop", "Keep", and "Rename" are processed at the end of the Data step and apply to the output dataset. "Where" is processed at the beginning of the Data step and applies to the output dataset.

Note that the syntax for "Rename" and "Where" require additional parentheses.

Practice

Use the scores1.sas dataset. Create a SAS dataset named scores1_copy, which is a duplicate of scores1.

Create a SAS dataset named scores1_mod which only includes the variables School and Score. Change the name of Score to be Final_Score. Only include records where the school is MET.

Debug the code below. The result should only include the Creatures variable.

```sas
data    occupancy;
Creatures = Residents + Dogs ;
set        data1.occupancy (drop=residents smokers
                                    Dogbreed
dogs);
run;
```

Readings

- Textbook sections 1.4, 6.1, 6.11

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder