MET MA 603: SAS Programming and Applications

Combining Datasets with Merge

Combining Datasets with Merge

The **Merge** Statement can be used to combine two or more SAS datasets. Usually, the datasets will have at least one variable in common. The BY statement is used to indicate the common variables used to merge the datasets. The datasets must be sorted by the BY variables.

The number of observations and the number of variables in the output dataset are the union of the observations and variables in the input datasets, respectively.

If a variable is present in more than one dataset, only the values from the first dataset with the variable are written for those observations.

There are two ways that datasets can be combined with the Merge statement: One-to-One and One-to-Many.

One-to-One Merge

In a **One-to-One Merge**, the input datasets all have the same number of observations. A One-to-One merge is used when the variables we want to work with are divided among more than one dataset.


```
Data   cities_with_pop ;
Merge  cities
                  city_populations ;
By                city ;
run ;
```


In a One-to-One merge, the number of observations in output and input datasets are the same.

One-to-Many Merge

In a **One-to-Many Merge**, the input datasets have different numbers of observations. In a one-to-many merge several variables in one input dataset correspond to a single variable in another input dataset.


```
Data   cities_with_pop_and_region ;
Merge  cities_with_pop
       regions ;
By     state ;
run ;
```


In a One-to-Many merge, the number of observations in output and input datasets are different.

The distinction between One-to-Many and One-to-One merges is made only to help in understanding the concept of merging – the coding looks exactly the same.

Practice

Merge the State_Info.sas7bdat and Policies_by_State.sas7bdat datasets according to the common variables.

Practice

The Policy_Info.sas7bdat dataset has a list of policies and the type of roof for the home. The Roof_Rating.sas7bdat dataset has the rating factor for each type of roof. Merge the datasets so that the correct rating factor for each policy is shown. Your result should match what is below (note all of the observations are shown).

VIEWTABLE: Work.Policy_roof_factors

	PolicyNumber	Material	State	Factor
1	3001	wood shake	DE	1.10
2	3002	wood shake	MS	1.10
3	3003	concrete tile	NC	0.95
4	3004	other	MS	1.00
5	3005	asphalt	SD	1.00
6	3006	asphalt	TX	1.00
7	3007	wood shingles	WV	1.10
8	3008	concrete tile	ID	0.95
9	3009	wood shake	MS	1.10
10	3010	wood shingles	FL	1.10

Readings

- Textbook sections 6.4, 6.5, 6.6