

Chapter 121

Merging Two Datasets

Introduction

Occasionally, it is useful to merge two datasets according to the value of one or more common (index) columns. This module allows you to merge two datasets, or, alternatively, update one dataset with the contents of another.

For example, consider the following dataset, called County, which contains county-level information for two states.

County Dataset

State	County	Pop	Age	Income
TX	1	72	34	65
TX	2	33	42	45
TX	5	25	23	46
TX	6	54	36	65
TX	7	11	42	53
TX	8	28	25	62
TX	9	82	35	66
TX	10	5	40	75
TX	11	61	27	22
MD	2	5	23	69
MD	4	98	25	73
MD	3	64	29	75
MD	2	36	24	65
MD	1	24	25	66
MD	5	34	31	78
MD	6	89	22	81
MD	8	21	25	73
MD	7	21	30	62

A second dataset, called State, contains similar information at the state level.

State Dataset

State	Pop	Age	Income	Education
TX	23543	32	54	10.2
MD	10343	29	69	10.3
IN	5231	41	35	10.1
CA	29587	35	67	10.4
NY	18142	34	78	10.2

Merging Two Datasets

Suppose that we wish to update the county dataset with the corresponding information from the state dataset. The resulting dataset, called CountyState, might appear as follows.

CountyState Dataset

State	County	Pop	Age	Income	St Pop	St Age	St Income	St Education
TX	1	72	34	65	23543	32	54	10.2
TX	2	33	42	45	23543	32	54	10.2
TX	5	25	23	46	23543	32	54	10.2
TX	6	54	36	65	23543	32	54	10.2
TX	7	11	42	53	23543	32	54	10.2
TX	8	28	25	62	23543	32	54	10.2
TX	9	82	35	66	23543	32	54	10.2
TX	10	5	40	75	23543	32	54	10.2
TX	11	61	27	22	23543	32	54	10.2
MD	2	5	23	69	10343	29	69	10.3
MD	4	98	25	73	10343	29	69	10.3
MD	3	64	29	75	10343	29	69	10.3
MD	2	36	24	65	10343	29	69	10.3
MD	1	24	25	66	10343	29	69	10.3
MD	5	34	31	78	10343	29	69	10.3
MD	6	89	22	81	10343	29	69	10.3
MD	8	21	25	73	10343	29	69	10.3
MD	7	21	30	62	10343	29	69	10.3

Only those states from the State dataset that were included on the County dataset were transferred to the resulting CountyState dataset.

Missing Values

The basic principle governing the treatment of missing values is that they cannot be matches. That is, even though values for corresponding by columns are both blank (missing), they will not be considered as a match. Only matches of non-missing values are recognized by the procedure.

Example 1 – Merging Two Datasets

This section presents an example of how to merge the two datasets, County and State, shown in the example above. The %P% and the %mydocs_NCSS% tags will be replaced by appropriate folders.

Setup

To run this example, complete the following steps:

1 Specify the Merge Two Datasets procedure options

- Find and open the **Merge Two Datasets** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

```

Merge Tab
-----
Merge (A).....%P%\Example Data\County.NCSS
With (B).....%P%\Example Data\State.NCSS
To Make.....%mydocs_NCSS%\Data\CountyState.NCSS
Match this Column from Dataset A .....State
With this Column from Dataset B.....State
Copy these Columns from A.....County-Income
Keep All Rows in this Dataset (Dataset A).....Checked
Copy these Columns from B.....Pop-Education
Append to Names (Dataset B)....._Total
    
```

2 Run the procedure

- Click the **Run** button to merge the datasets.

This procedure does not produce any output. Instead, the resulting dataset is opened in the data table.