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

© NCSS, LLC. All Rights Reserved.

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

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.

© NCSS, LLC. All Rights Reserved.

Procedure Options

This section describes the options available in this procedure.

Merge Tab

This panel specifies the datasets and columns to be merged.

Datasets

Merge (A)

This is the fully-qualified name of the first dataset. The contents of the second dataset will be merged with this one according to the values of the corresponding By Columns.

With (B)

This is the fully-qualified name of the second dataset. The contents of this dataset are merged with dataset A. Only rows in this dataset that have matching values in the corresponding 'By Columns' will be kept.

To Make

This is the fully-qualified name of the dataset produced by the merge operation. '*Fully-qualified*' means that the drive and folder containing the dataset are included in the entry. If no name is entered, the resulting dataset will be "Untitled" and can be saved later.

Existing data in this dataset will be replaced, so do not select a dataset that contains data you need to keep.

Merge By Matching Values from the Following Column Pairs

Match this Column from Dataset A

Specify a single column from dataset A whose values will be used by comparing them with those of the corresponding column from dataset B. Note that blanks in both values are not considered to be a match!

With this Column from Dataset B

Specify a single column from dataset B whose values will be used by comparing them with those of the corresponding column from dataset A. Note that blanks in both values are not considered to be a match!

Options for Datasets A and B

Copy these Columns from A (or B)

Select the columns from dataset A (or B) that are to be retained in the resulting dataset. Note that this does not include the match columns, since they are included automatically.

Prepend to Names (Datasets A & B)

Specify a few letters to be added to the beginning of each column name that was kept from the corresponding dataset (A or B). This allows you to rename columns when there are names that are common to both datasets that might cause confusion in the resulting dataset.

If you do not want to change the column names of a dataset, leave this option blank.

For example, if the column names that were kept were 'TIME' and 'AMOUNT', and 'AA_' is specified here, the resulting column names are 'AA_TIME' and 'AA_AMOUNT'.

Note the only letters, integers, and the underscore may be added.

© NCSS, LLC. All Rights Reserved.

Append to Names (Datasets A & B)

Specify a few letters to be added to the end of each column name that was kept from the corresponding dataset (A or B). This allows you to rename columns when there are names that are common to both datasets that might cause confusion in the resulting dataset.

If you do not want to change the column names of a dataset, leave this option blank.

For example, if the column names that were kept were 'TIME' and 'AMOUNT', and '_TOTAL' was specified here, the resulting column names would be 'TIME_TOTAL' and 'AMOUNT_TOTAL'.

Note the only letters, integers, and the underscore may be added.

Keep All Rows in this Dataset (A)

Check this box if you want to keep all rows from dataset A in the resulting dataset, even if they do not have a match in dataset B.

If this box is not checked, rows with missing values in the 'By Column' and rows that do not have a match in dataset B will be omitted from the resulting dataset.

Keep All Rows in this Dataset (B)

Check this box if you want to keep all rows from dataset B in the resulting dataset.

If this box is not checked, rows with missing values in the 'By Column' and rows that do not have a match in dataset A will be omitted from the resulting dataset.

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.

You may follow along here by making the appropriate entries or load the completed template **Example – Merging Two Datasets** by clicking on Open Example Template from the File menu of the Merging Two Datasets window.

1 Open the Merge Two Datasets window.

- Using the Data menu or the Procedure Navigator, find and select the **Merging Two Datasets** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

2 Specify the datasets.

- Specify the **Merge (A) Dataset** as **%P%\Example Data\County.NCSS**.
- Specify the **With (B) Dataset** as **%P%\Example Data\State.NCSS**.
- Specify the **To Make Dataset** as **%mydocs_NCSS%\Data\CountyState.NCSS**.
- Set **Match this Column from Dataset A** as **State**.
- Set **With this Column from Dataset B** as **State**.
- Set **Copy these Columns from A** as **County-Income**.
- Set **Copy these Columns from B** as **Pop-Education**.
- Check the **Keep All Rows in Dataset for Dataset A** option.
- Set **Append to Names for dataset B** as "**_Total**" (no quotes).

3 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

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

© NCSS, LLC. All Rights Reserved.