Is it possible to merge two sets, A and B, into a single set by taking only the elements that are in A but not in B?

Authored by

stats writer

June 23, 2024

The concept of merging two sets, A and B, into a single set by taking only the elements that are in A but not in B, refers to the process of combining two distinct sets while removing any duplicate elements. This results in a new set that contains all the unique elements from both sets A and B. This operation is known as set difference or set complement and is applicable when dealing with sets in mathematics and computer science. It allows for the creation of a new set that contains only the elements that are present in one set but not in the other, making it a useful tool for various applications.

SAS: Merge If A Not B

You can use the following basic syntax to merge two datasets in SAS and only return the rows where a value exists in the first dataset and not the second dataset:

```
data final_data;
merge data1 (in = a) data2 (in = b);
by ID;
if a and not b;
run;
```

This particular example merges the datasets called data1 and data2 and only returns the rows where a value exists in data1 and not in data2.

The following example shows how to use this syntax in practice.

Example: Merge Two Datasets in SAS Using If A Not B

Suppose we have the following two datasets in SAS that contain information about sales associates at some company:

```
/*create first dataset*/
data data1;
input ID Gender $;
datalines;
1 Male
2 Male
3 Female
4 Male
5 Female
;
run;

title "data1";
proc printdata = data1;

/*create second dataset*/
data data2;
input ID Sales;
datalines;
```

```
1 22
2 15
4 29
6 31
7 20
8 13
;
run;

title "data2";
proc printdata = data2;
```

data1

Obs	ID	Gender
1	1	Male
2	2	Male
3	3	Female
4	4	Male
5	5	Female

data2

Obs	ID	Sales
1	1	22
2	2	15
3	4	29
4	6	31
5	7	20
6	8	13

If we use the following merge statement to merge the two datasets based on the value in the ID column, all rows will be returned from both datasets:

```
/*perform merge*/
data final_data;
merge data1 data2;
by ID;
run;
```

```
/*view results*/
title "final_data";
proc printdata=final_data;
```

final_data

Obs	ID	Gender	Sales
1	1	Male	22
2	2	Male	15
3	3	Female	.
4	4	Male	29
5	5	Female	.
6	6		31
7	7		20
8	8		13

However, we can use IN statements to only return the rows where a value exists in the first dataset and *not* in the second dataset:

```
data final_data;
merge data1 (in = a) data2 (in = b);
by ID;
if a and not b;
run;

/*view results*/
```

```
title "final_data";
proc printdata=final_data;
```

final_data

Obs	ID	Gender	Sales
1	3	Female	.
2	5	Female	.

Notice that only the rows where a value exists in the first dataset and not the second dataset are returned.

Note: You can find the complete documentation for the SAS merge statement .