

How can I perform a left join in SAS, and could you provide an example?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform a left join in SAS, and could you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164282>

A left join in SAS is a method used to combine two datasets based on a common variable, where all the observations from the left dataset are retained, and only matching observations from the right dataset are included. This can be achieved by using the MERGE statement in SAS, specifying the left dataset first, followed by the right dataset, and then using the keyword "LEFT" after the MERGE statement. An example of this would be merging a dataset of customer information with a dataset of product purchases, using the customer ID as the common variable, to get a list of all customers and their corresponding purchases.

Perform a Left Join in SAS (With Example)

You can use the following basic syntax to perform a left join with two datasets in SAS:

```
proc sql;  
createtable final_table as  
select * from data1 as x left join data2 as y  
on x.ID = y.ID;  
quit;
```

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

Related:

Example: Left Join in SAS

Suppose we have the following two datasets in SAS:

```
/*create datasets*/
```

```
data data1;  
input team $ points;  
datalines;  
Mavs 99  
Spurs 93  
Rockets 88  
Thunder 91  
Warriors 104  
Cavs 93  
Grizzlies 90  
Hawks 91  
;  
run;  
  
data data2;  
input team $ rebounds;  
datalines;  
Mavs 21  
Spurs 18  
Rockets 22  
Warriors 27  
Cavs 15  
Hawks 29  
;
```

```
run;
```

```
/*view datasets*/
```

```
proc printdata=data1;
```

```
proc printdata=data2;
```

Obs	team	points
1	Mavs	99
2	Spurs	93
3	Rockets	88
4	Thunder	91
5	Warriors	104
6	Cavs	93
7	Grizzlie	90
8	Hawks	91

Obs	team	rebounds
1	Mavs	21
2	Spurs	18
3	Rockets	22
4	Warriors	27
5	Cavs	15
6	Hawks	29

Notice that the two datasets share one variable in common: team.

We will use the following syntax to perform a left join

and create a new dataset that contains every row from data1 and *only* the rows from data2 that match a team name in data1:

```
/*perform left join*/  
proc sql;  
createtable final_table asselect * from data1 as x leftjoin  
data2 as y  
on x.team = y.team;  
quit;/*view results of left join*/  
proc print data=final_table;
```

Obs	team	points	rebounds
1	Cavs	93	15
2	Grizzlies	90	.
3	Hawks	91	29
4	Mavs	99	21
5	Rockets	88	22
6	Spurs	93	18
7	Thunder	91	.
8	Warriors	104	27

The resulting dataset contains every original team from data1, but the only teams that have values for the rebounds column are the ones that also appeared in data2.

Additional Resources

The following tutorials explain how to perform other common tasks in SAS:

ARABPSYCHOLOGY.COM