

How can I use Proc Compare in SAS, and can you provide some examples?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use Proc Compare in SAS, and can you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160498>

Proc Compare is a procedure in SAS that allows users to compare two datasets and identify any differences between them. It is useful for quality control, data validation, and identifying errors in data. To use Proc Compare, users can specify the two datasets to be compared, as well as the variables and criteria for comparison. The procedure will then generate a report showing the differences between the two datasets. Some examples of using Proc Compare include comparing data before and after a data transformation, verifying data from different sources, and identifying data discrepancies between production and test environments.

Use Proc Compare in SAS (With Examples)

You can use PROC COMPARE in SAS to quickly identify the similarities and differences between two datasets.

This procedure uses the following basic syntax:

```
proc compare base=data1  
compare=data2;  
run;
```

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

Example: Using Proc Compare in SAS

Suppose we have the following two datasets in SAS:

```
/*create datasets*/  
data data1;  
input team $ points rebounds;
```

```
datalines;
```

```
A 25 10
```

```
B 18 4
```

```
C 18 7
```

```
D 24 12
```

```
E 27 11
```

```
;
```

```
run;
```

```
data data2;
```

```
input team $ points;
```

```
datalines;
```

```
A 25
```

```
B 18
```

```
F 27
```

```
G 21
```

```
H 20
```

```
;
```

```
run;
```

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

```
proc printdata=data1;
```

```
proc printdata=data2;
```

Obs	team	points	rebounds
1	A	25	10
2	B	18	4
3	C	18	7
4	D	24	12
5	E	27	11

Obs	team	points
1	A	25
2	B	18
3	F	27
4	G	21
5	H	20

We can use the following PROC COMPARE statement to find the similarities and differences between the two datasets:

```
/*compare the two datasets*/  
proc comparebase=data1  
compare=data2;  
run;
```

This will produce three tables in the output:

Table 1: A Summary of Both Tables

The COMPARE Procedure
Comparison of WORK.DATA1 with WORK.DATA2
(Method=EXACT)

Data Set Summary

Dataset	Created	Modified	NVar	NObs
WORK.DATA1	07MAR22:14:31:17	07MAR22:14:31:17	3	5
WORK.DATA2	07MAR22:14:31:17	07MAR22:14:31:17	2	5

Variables Summary

Number of Variables in Common: 2.

Number of Variables in WORK.DATA1 but not in WORK.DATA2: 1.

The first table shows a brief summary of each dataset, including:

1. The number of variables (NVar) and observations (NObs) in each dataset.

Data1 has 3 variables and 5 observations Data2 has 2 variables and 5 observations

2. The number of variables in common between the two datasets.

Data1 and Data2 have 2 variables in common (team and points)

Table 2: A Summary of the Number of Differences in

Values

Observation Summary		
Observation	Base	Compare
First Obs	1	1
First Unequal	3	3
Last Unequal	5	5
Last Obs	5	5

Number of Observations in Common: 5.
 Total Number of Observations Read from WORK.DATA1: 5.
 Total Number of Observations Read from WORK.DATA2: 5.

 Number of Observations with Some Compared Variables Unequal: 3.
 Number of Observations with All Compared Variables Equal: 2.

Values Comparison Summary		
Number of Variables Compared with All Observations Equal: 0.		
Number of Variables Compared with Some Observations Unequal: 2.		
Total Number of Values which Compare Unequal: 6.		
Maximum Difference: 9.		

All Variables Compared have Unequal Values

Variable	Type	Len	Ndif	MaxDif
team	CHAR	8	3	
points	NUM	8	3	9.000

The second table summarizes the number of differences in values between the two tables.

The most interesting part of this output is located at the end of the table where we can see a summary of differences between the variables:

The team variable has 3 observations with different values. The points variables has 3 observations with different values. The max difference is 9.

Table 3: The Actual Differences Between Observations

The COMPARE Procedure					
Comparison of WORK.DATA1 with WORK.DATA2					
(Method=EXACT)					
Value Comparison Results for Variables					
Obs		Base Value	Compare Value		
		team	team		
3		C	F		
4		D	G		
5		E	H		
Obs		Base points	Compare points	Diff.	% Diff
3		18.0000	27.0000	9.0000	50.0000
4		24.0000	21.0000	-3.0000	-12.5000
5		27.0000	20.0000	-7.0000	-25.9259

The third table shows the actual differences between the observations in the two datasets.

The first table shows the differences in the team variable between the two datasets.

For example, in data1 the third observation has a value of C for team while in data2 the third observation has a value of F.

The second table shows the differences in the points variables between the two datasets.

For example, in data1 the third observation has a value of 18 for points while in data2 the third observation has a value of 27. The difference between the two values is 9.

These three tables give us a complete understanding of the differences between the two datasets.

Note that if you only want to compare the differences between the two datasets for one specific variable, you can use the following syntax:

```
/*compare the differences between the datasets only for  
'points' variable*/  
proc comparebase=data1compare=data2;  
var points;run;
```

This will produce the same three tables as earlier, but

only the output for the points variable will be shown.

Note: You can find the complete documentation for PROC COMPARE .

Additional Resources

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

ARABPSYCHOLOGY.COM