

# How can I remove duplicates in SAS, and what are some examples of this process?

Authored by  
**stats writer**

July 1, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I remove duplicates in SAS, and what are some examples of this process?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164984>

Removing duplicates in SAS is a process of eliminating repeated observations in a dataset. This can be achieved by using the SORT procedure followed by either the BY or NODUPKEY statement. The SORT procedure arranges the data in a specified order, while the BY statement groups the data by a particular variable and removes duplicates within each group. Similarly, the NODUPKEY statement removes duplicates based on a specific variable and keeps only the first observation for each unique value. Some examples of removing duplicates in SAS include cleaning data for analysis, merging datasets, and creating unique identifier variables.

## Remove Duplicates in SAS (With Examples)

You can use `proc sort` in SAS to quickly remove duplicate rows from a dataset.

This procedure uses the following basic syntax:

```
proc sort data=original_data out=no_dups_data
nodupkey;
by _all_;
run;
```

Note that the `by` argument specifies which columns to analyze when removing duplicates.

The following examples show how to remove duplicates from the following dataset in SAS:

```
/*create dataset*/
data original_data;
```

```
input team $ position $ points;  
datalines;  
A Guard 12  
A Guard 20  
A Guard 20  
A Guard 24  
A Forward 15  
A Forward 15  
A Forward 19  
A Forward 28  
B Guard 10  
B Guard 12  
B Guard 12  
B Guard 26  
B Forward 10  
B Forward 10  
B Forward 10  
B Forward 19  
;  
run;  
  
/*view dataset*/  
proc printdata=original_data;
```

Obs	team	position	points
1	A	Guard	12
2	A	Guard	20
3	A	Guard	20
4	A	Guard	24
5	A	Forward	15
6	A	Forward	15
7	A	Forward	19
8	A	Forward	28
9	B	Guard	10
10	B	Guard	12
11	B	Guard	12
12	B	Guard	26
13	B	Forward	10
14	B	Forward	10
15	B	Forward	10
16	B	Forward	19

### Example 1: Remove Duplicates from All Columns

We can use the following code to remove rows that have duplicate values across all columns of the dataset:

```
/*create dataset with no duplicate rows*/
```

```
proc sort data=original_data out=no_dups_data  
nodupkey;
```

```
by _all_;
```

```
run;
```

```
/*view dataset with no duplicate rows*/
```

```
proc printdata=no_dups_data;
```

Obs	team	position	points
1	A	Forward	15
2	A	Forward	19
3	A	Forward	28
4	A	Guard	12
5	A	Guard	20
6	A	Guard	24
7	B	Forward	10
8	B	Forward	19
9	B	Guard	10
10	B	Guard	12
11	B	Guard	26

**Notice that a total of five duplicate rows have been removed from the original dataset.**

**Example 2: Remove Duplicates from Specific Columns**

**We can use the by argument to specify which columns to look at when removing duplicates.**

**For example, the following code removes rows that have duplicate values in the team and position columns:**

```
/*create dataset with no duplicate rows in team and position columns*/
```

```
proc sortdata=original_data out=no_dups_data  
nodupkey;  
by team position;  
run;
```

```
/*view dataset with no duplicate rows in team and  
position columns*/  
proc printdata=no_dups_data;
```

Obs	team	position	points
1	A	Forward	15
2	A	Guard	12
3	B	Forward	10
4	B	Guard	10

### Additional Resources

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