**SPSS CODE FRAGMENT GENERATING ID NUMBERS WITHIN GROUP OR PANEL**

The code below shows how to create an id variable that takes on the values 1 to *n* within each group/panel (where *n* is the number of cases in a group).  The grouping variable is defined by the user.  The number of cases, *n*, need not be the same across all groups.  In the example dataset, **group** is the grouping variable.  The variables **v1** and **v2** are additional variables in the dataset and are not modified by the code.

**data list list /group v1 v2.**

**begin data**

**1 12 0**

**1 23 1**

**1 22 0**

**3 5 0**

**3 7 0**

**3 13 1**

**5 1 0**

**5 56 1**

**end data.**

**sort cases by group.**

**compute id = 1.**

**if group = lag(group) id = lag(id) + 1.**

**exe.**

**list group id v1 v2.**

group id v1 v2

1.00 1.00 12.00 .00

1.00 2.00 23.00 1.00

1.00 3.00 22.00 .00

3.00 1.00 5.00 .00

3.00 2.00 7.00 .00

3.00 3.00 13.00 1.00

5.00 1.00 1.00 .00

5.00 2.00 56.00 1.00

Number of cases read: 8 Number of cases listed: 8

### SORTING WITH SYNTAX

This syntax performs the same sort shown in the screencap above (sort by ascending class rank, then descending birth date).

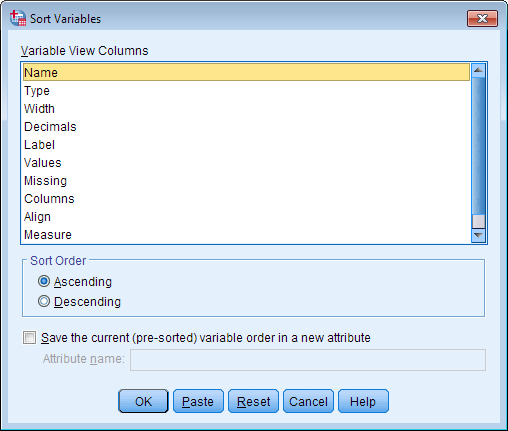
**SORT CASES BY Rank(A) bday(D).**

## Sort Variables

Sorting variables will rearrange the order of the variables (columns) in your data. Variables can be sorted on only one attribute at a time: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, or a custom attribute. Variables can be sorted in ascending or descending order with respect to the selected attribute.

To sort variables, follow these steps:

1. Click **Data > Sort Variables**.
2. The Variable View Columns list includes the attributes of variables that may be used to sort variables, including: name, type, width, decimals, label, values, missing, columns, align, measure. Select an attribute by clicking it in the list, which will highlight your selection. Note that you can only select one variable attribute. In this example, “Name” has been selected, which means that variables will be sorted according to their names.

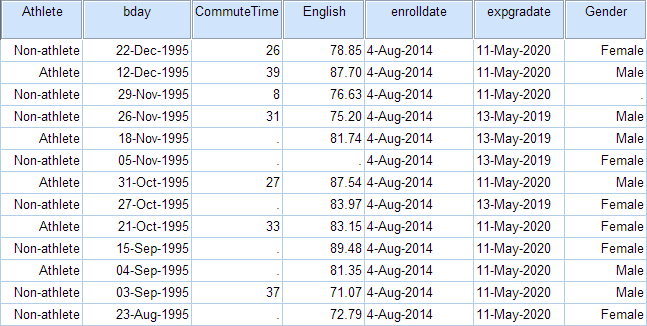


1. In the Sort Order area, variables can be sorted in “Ascending” or “Descending” order. Click the radio button that corresponds to your choice. In this example, variables will be sorted in ascending order.

You may also choose to save the current (pre-sorted) variable order in a new attribute by selecting the Attribute name check box and typing a name into the text field. Custom attributes are simply user-defined characteristics of variables that can be used to sort variables.

1. When you are finished, click **OK**.

Now your variables will be sorted according to the attribute you selected. In this example, the variables are sorted in ascending order according to their names (i.e., alphabetically).



### SYNTAX

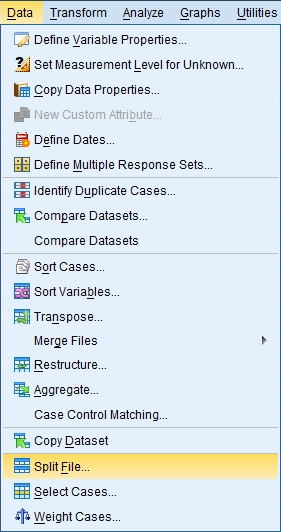
**SORT VARIABLES BY NAME (A).**

## Grouping or Splitting Data

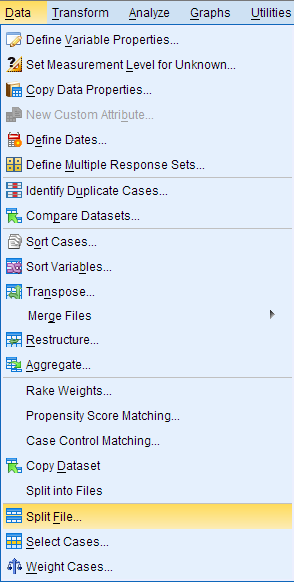
When analyzing data, it is sometimes useful to temporarily "group" or "split" your data in order to compare results across different subsets. This can be useful when you want to compare frequency distributions or descriptive statistics with respect to the categories of some variable (e.g., Gender) - especially if you want separate tables of results for each group.

To split your dataset, click **Data > Split File**.

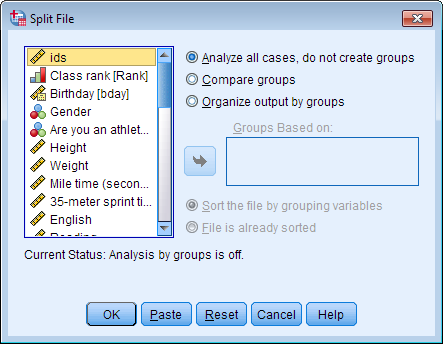
SPSS Version 21 Drop-Down Menu



SPSS Version 22 Drop-Down Menu



The Split File window will appear. By default, the dataset is not split according to any criteria; this is indicated by **Analyze all cases, do not create groups**.



You can choose one of two ways to split the data:

1. Compare groups
2. Organize output by groups

For both splitting methods, there are two considerations to be made:

* The splitting variable(s) should be nominal or ordinal categorical. SPSS will not stop you from using a continuous variable as a splitting variable, but it is a bad idea to try to attempt this; SPSS will see each unique numeric value as a distinct category.
* In order to split the file, SPSS requires that the data be sorted with respect to the splitting variable. By default, **Sort the file by grouping variables** is selected.

## Turning Off Split File

When you no longer want to split your analyses by group, you can turn Split File off through the same window you used to turn it on.

1. Click **Data > Split File.**
2. Click **Analyze all cases, do not create groups**.
3. Click **OK**.

You can now run all analyses normally again.

### SYNTAX

**SPLIT FILE OFF.**

## Example

What are the differences in the split file options?

The Compare and Organize options produce numerically identical results when the same grouping variable(s) are applied. This is true regardless of what statistical analysis is used. The difference between the two options is how the numeric results are presented.

* If **Compare groups**is used, then all of the results will be shown in a single table. The table will have sections showing the results for each group.
* If **Organize output by groups**is used, then each groups' results will be put into a separate table.

The choice of which splitting method to use is entirely about what format the user wants their results in. Do you want a single table with all results, or separate tables for each group's results? A good rule of thumb is to choose Compare Groups if you want to be able to directly compare the results of your groups, and to choose Organize Output by Groups if the information is from separate trials or samples (such as cohorts from different years).

### PROBLEM STATEMENT

Suppose that we want to get a summary of the differences in height between males and females in the sample data. Let's couple the Split File procedure with the Descriptives procedure to get summary statistics for the two groups. We'll use both Split File methods so that we can compare what their outputs look like.

### SPLITTING USING COMPARE GROUPS

If you choose to split your data using the **Compare groups** option and then run a statistical analysis in SPSS, your output will be displayed in a single table that organizes the results according to the grouping variable(s) you specified.

#### RUNNING THE PROCEDURE

To split the data in a way that will facilitate group comparisons:

1. Click **Data > Split File**.
2. Select the option **Compare groups**.
3. Double-click the variable Gender to move it to the **Groups Based on**field.
4. When you are finished, click **OK**.

After splitting the file, the only change you will see in the Data View is that data will be sorted in ascending order by the grouping variable(s) you selected.

Now let's view the aforementioned descriptive statistics for the variable Height with respect to Gender. Select **Analyze > Descriptive Statistics > Descriptives**. Double click on the Height variable, then click **OK**.

##### **Syntax**

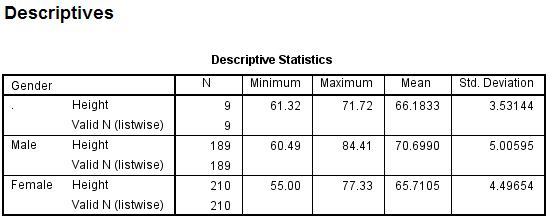
**SORT CASES BY Gender.**

**SPLIT FILE LAYERED BY Gender.**

**DESCRIPTIVES VARIABLES=Height**

**/STATISTICS=MEAN STDDEV MIN MAX.**

#### OUTPUT



This table gives us a breakdown of how many observations were in each group (N), and the minimum, maximum, average, and standard deviation of each group. The '.' group contains cases with missing gender values and nonmissing height values. At a glance, we can quickly take note that in this sample:

* The height of the tallest male was greater than the height of the tallest female.
* The male heights tended to have a slightly larger standard deviation (spread) than the female heights.
* On average, the males were taller than the females.
* The individuals with missing values for gender had a much smaller range of heights than did the males or females.

Note: This combination of Split File: Compare Groups with Descriptives is very similar to what you would get with the Compare Means procedure. The major difference is that Split File includes the missing values in the grouping/splitting variable, whereas Compare Means excludes missing values in the grouping variable.

### SPLITTING USING ORGANIZE OUTPUT BY GROUPS

If you choose to split your data using the **Organize output by groups** option and then run a statistical analysis in SPSS, your output will be broken into separate tables for each category of the grouping variable(s) specified.

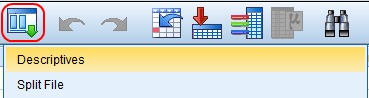
#### RUNNING THE PROCEDURE

To split the data in a way that separates the output for each group:

1. Click **Data > Split File**.
2. Select the option **Organize output by groups**.
3. Double-click the variable Gender to move it to the **Groups Based on**field.
4. When you are finished, click **OK**.

After splitting the file, the only change you will see in the Data View is that data will be sorted in ascending order by the grouping variable(s) you selected.

Now we will re-run the same descriptive statistics procedure that we ran before. You can go through the menu system again (**Analyze > Descriptive Statistics > Descriptives**), or you can click on the **Recall recently used dialogs** icon, which will bring up a list of recently used procedures:



##### **Syntax**

**SORT CASES BY Gender.**

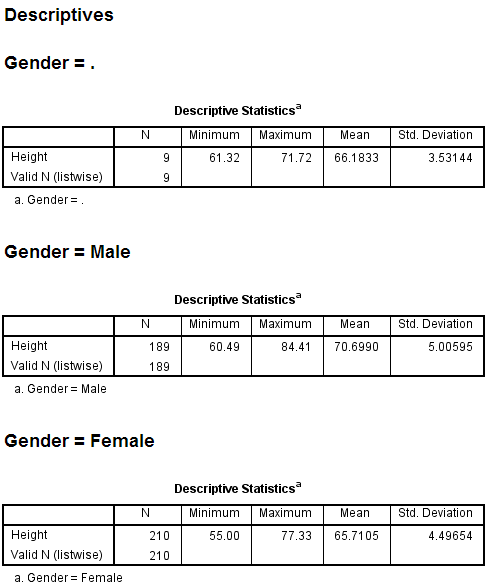
**SPLIT FILE SEPARATE BY Gender.**

**DESCRIPTIVES VARIABLES=Height**

**/STATISTICS=MEAN STDDEV MIN MAX.**

#### OUTPUT

After re-running the descriptive statistics, we see that the output is broken into three sections based on values of the Gender variable. The first section (“Gender = .”) reports the minimum, maximum, average, and standard deviation of Height for the students who had missing values for Gender. The second section reports those same statistics for the male students; the third section reports the statistics for the females.



# ARCHIVED: In SPSS, how do I drop cases with missing values?

The SELECT command with the SYSMIS() function can drop all missing cases from the current SPSS data set. Consider the following:

SELECT IF NOT (SYSMIS(amount)).

SAVE OUTFILE= 'newfile.sav'.

This example drops all cases whose value of the variable amount is missing, and then saves this data to an SPSS system file called newfile.sav.

If the data set has more than one coding for missing values, as is often the case for survey data, select all of the different codings for missing values with the AND operator:

SELECT IF NOT (SYSMIS(amount1)) AND NOT (SYSMIS(amount2)).

SAVE OUTFILE= 'newfile.sav'.

## SPSS IF Syntax Example 1

**\*1. Create test data.**  
data list free/gender score.  
begin data  
0 80 1 85 0 90 1 95 0 '' 1 105 0 110 1 115  
end data.  
  
**\*2. Replace missing value with 100.**  
if missing(score) score = 100.  
  
**\*3. Sort cases.**  
sort cases gender.

## SPSS IF Syntax Example 2

**\*1. Create score groups option 1.**  
if score lt 100 group\_a = 1.  
if score ge 100 group\_a = 2.  
exe.  
  
**\*2. Create score groups option 2.**  
recode score (100 thru hi = 2) (else = 1) into group\_b.  
exe.  
  
**\*3. Create score groups option 3.**  
compute group\_c = (score ge 100) + 1.  
exe.

## SPSS IF Syntax Example 3

**\*1. Gender-score groups option 1.**  
if score lt 100 and gender eq 0 group\_d = 1.  
if score ge 100 and gender eq 0 group\_d = 2.  
if score lt 100 and gender eq 1 group\_d = 3.  
if score ge 100 and gender eq 1 group\_d = 4.  
exe.  
  
**\*2. Gender-score groups option 2.**  
compute group\_e = 2 \* gender + (score ge 100) + 1.  
exe.