


28 Analysing Two or more Groups - Split Files

If you want to analyse your data by sub groups within the study sample, Split Files is the command to use. Suppose you want to analyse your data by gender groups, male and female separately.

In the Data Editor window,

Either **Click** on the Split file button 
or Select **Data | Split File** from the menu.

Click on *Compare Groups*

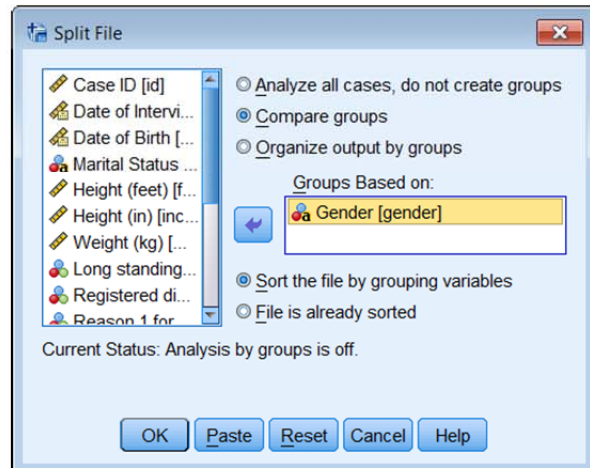
Highlight *gender* in the variable list drag it or press  to add this to the *Groups based on:* panel.

Check that the *Sort the file by grouping variables* option is selected.

Click on **Paste**.

In the syntax window,
highlight the commands

SORT CASES BY gender .
SPLIT FILE LAYERED BY gender .



Click on the **Run Selection** button .

All further analyses will be carried out for females and males separately, rather than for all cases.

The difference between the output produced by the options *Compare Groups* and *Organise output by groups* is illustrated below:

Compare Groups

	Frequency
Females	
Males	

Organise output by groups

Gender =
Female

Frequency

Gender = Male

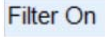
Frequency

Organise output by groups creates more lines of output although the same statistics are produced.

Note: the SPSS status bar changes to show that Split Files is active, **Split by sex** is displayed.

To turn the split files effect off, select **Data | Split Files** or click on  and **Click** on *Analyze all cases, do not create groups*. **Click** on **OK**, or **Paste** the command SPLIT FILE OFF into the syntax window and run it.

29 Case selection – Filtering

SPSS provides a 'filtering' option for analysing a subset of your cases. When filtering is turned on, as shown on the status bar , only those cases meeting the selection criterion are used for analysis. The criteria used to define sub groups can include:

- Variable values and ranges
- Date and time ranges
- Case (row) numbers
- Arithmetic expressions
- Logical expressions
- Functions

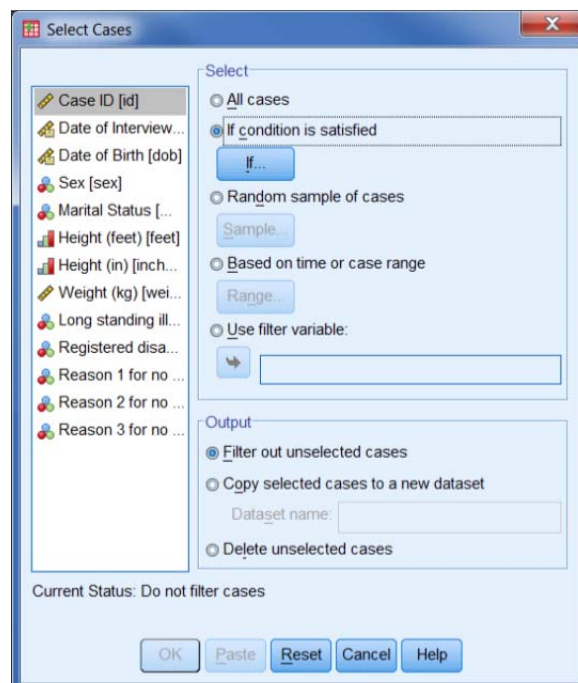
Suppose that you wished to analyse the data for those respondents that have a long standing illness (the variable LSILL will have a value of 1). In the Data Editor window, either:

Click on the Select Cases button 

or

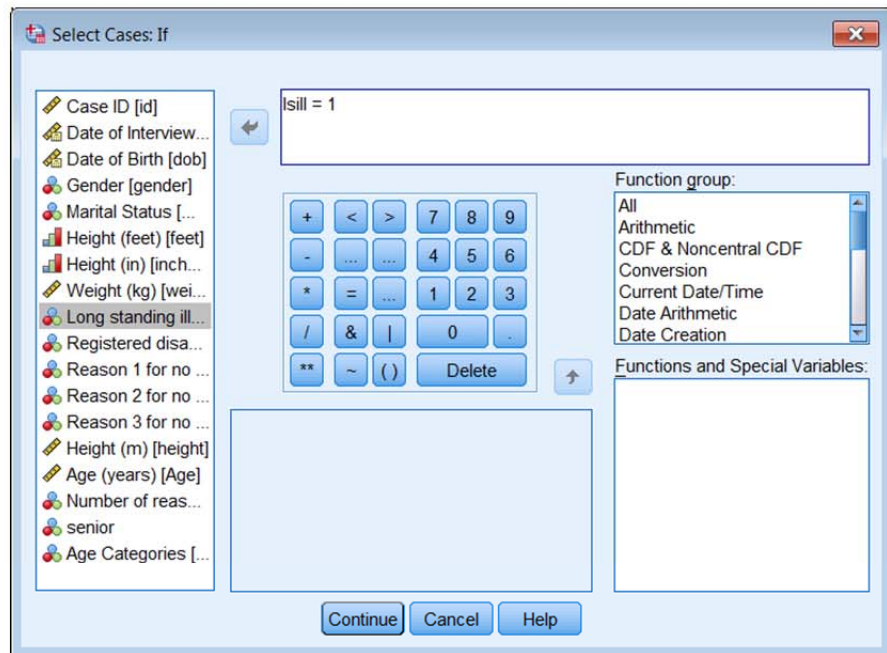
Select **Data | Select Cases** from the menu.

There are four ways of selecting cases, the *If condition is satisfied* option is the most frequently used.



Choose this option by clicking the **If...** button. A *Select Cases: If* dialogue box is displayed.

If we wish to analyse those cases where the variable value is 1, an expression **Isill = 1** is built in the text panel by clicking on the buttons.



Click on **Continue**.


Paste and **Run** the commands in the Syntax window

```
USE ALL.
COMPUTE filter_$=(Isill = 1).
VARIABLE LABEL filter_$ 'Isill = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
```

Filtered (out) cases are left in the data editor with a diagonal line through the case number. A variable, **filter_\$**, is also created, this has a value of 1 for cases that are Selected (10 and 11) and 0 for cases that are Not Selected (7, 8, 9, 12 and 13).

The Status bar in the Data Editor displays **Filter On**.

Case	id	doi	dob	gender	mstatus	feet	inches	weight	Isill	regdisab
5	69	03-Aug-20...	23-Mar-1975	Female	Single	5	4	77.0	No	Not Applic...
6	94	10-Jul-2004	13-Dec-1984	Female	Married	5	6	58.1	No	Not Applic...
7	143	24-Jul-2004	23-May-1938	Female	Widowed	5	2	68.5	No	Not Applic...
8	160	23-Jul-2004	19-Aug-1936	Female	Married	5	8	54.0	No	Not Applic...
9	186	15-Jul-2004	10-Jun-1981	Male	Single	6	2	76.2	Yes	No
10	211	09-Aug-20...	16-Jan-1958	Male	Married	6	0	80.7	Yes	No
11	224	11-Aug-20...	16-Jan-1925	Female	Widowed	5	2	89.8	No	Not Applic...
12	244	03-Jun-2004	02-Jan-1972	Male	Married	5	4	52.2	No	Not Applic...

To turn the Filtering off, click  or select **Data | Select Cases**.
Click on *All Cases*.

Paste and **Run** the commands.

FILTER OFF.
USE ALL.
EXECUTE .

If you want to perform a single analysis on a subgroup, use filtering. But if you want to do intensive work with selected cases, *Delete unselected cases* is probably more efficient. In this case it is worthwhile saving these cases to another SPSS data file.

30 Sorting Cases

The data set can be sorted by cases (rows) using **Data | Sort Cases**.

You can select multiple sort variables, cases are sorted by each variable within categories of the preceding variable on the *Sort by*: list. For example, if you select Gender as the first sorting variable and Weight as the second sorting variable, cases will be sorted by Weight within each Gender category.

The *Sort Order* can be Ascending or Descending for each variable.

