

Σ+ SPSS TUTORIALS

BASICS DATA ANALYSIS T-TEST ANOVA CHI-SQUARE TEST

SPSS FILTER – Quick & Simple Tutorial

SPSS FILTER temporarily excludes a selection of cases from all data analyses.

For excluding cases from data *editing*, use **DO IF** or **IF** instead.

- SPSS Filtering Basics
- Example 1 - Exclude Cases with Many Missing Values
- Example 2 - Filter on 2 Variables
- Example 3 - Filter without Filter Variable

You are here: [Home](#)

→ [Basics](#) → [Tip - Commands with Built-In Filters](#)

[Essential SPSS](#) • [Warning - Data Editing with Filter](#)

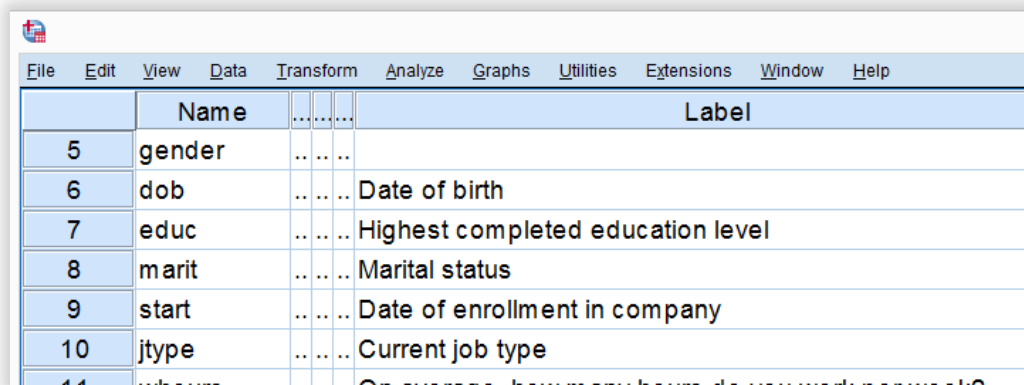
[Commands](#) → [SPSS](#)

[FILTER – Quick &](#)

[Simple Tutorial](#)

SPSS FILTER - Example Data

I'll use **bank_clean.sav** -partly shown below- for all examples in this tutorial. This file contains the data from a small bank employee survey. Feel free to download these data and rerun the examples yourself.

	Name	Label
5	gender	
6	dob	Date of birth
7	educ	Highest completed education level
8	marit	Marital status
9	start	Date of enrollment in company
10	jtype	Current job type
11	salary	On average, how many hours do you work per week?

SPSS Filtering Basics

Filtering in SPSS usually involves 4 steps:

AdChoices

Analyzing Data

SPSS Syntax

SPSS Data Analysis

1. **create** a filter variable;
2. **activate** the filter variable;
3. run one or many **analyses** -such as **correlations**, **ANOVA** or a **chi-square test**- with the filter variable in effect;
4. **deactivate** the filter variable.

In theory, *any* variable can be used as a filter variable. After activating it, cases with

Ne
Va

- zeroes,
- user missing values or
- system missing values

on the filter variable are **excluded** from all analyses until you deactivate the filter. For the sake of clarity, I recommend you

only use filter variables containing 0 or 1 for each case.

Enough theory. Let's put things into practice.

Example 1 - Exclude Cases with Many Missing Values

At the end of our data, we find 9 rating scales: q1 to q9. Perhaps we'd like to run a **factor analysis** on them or use them as predictors in **regression analysis**. In any case, we may want to exclude cases having many missing values on these variables. We'll first just count them by running the **syntax** below.

```
*Create new variable holding number of missings over  
compute mis_1 = nmiss(q1 to q9).
```

***Apply variable label.**

variable labels mis_1 'Number of missings on q1 to q9'

***Check frequencies.**

frequencies mis_1.

Result

Number of missings on q1 to q9

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	309	66.6	66.6	66.6
	1.00	125	26.9	26.9	93.5
	2.00	22	4.7	4.7	98.3
	3.00	3	.6	.6	98.9
	6.00	2			99.4
	8.00	1			99.6
	9.00	2	.4	.4	100.0
Total		464	100.0	100.0	

**8 CASES HAVE 3 OR MORE
MISSING VALUES ON Q1 TO Q9**

Based on this frequency distribution, we decided to exclude the 8 cases having 3 or more missing values on q1 to q9. We'll create our filter variable with a simple RECODE as shown below.

***Recode mis_1 into filter variable.**

recode mis_1 (10 thru 2 = 1) (else = 0) into filt_1.

***Apply variable label.**

variable labels filt_1 'Filter out cases with 3 or mo

***Activate filter variable.**

filter by filt_1.

***Reinspect numbers of missings over q1 to q9.**

frequencies mis_1.

Result

Number of missings on q1 to q9					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	309	67.8	67.8	67.8
	1.00	125	27.4	27.4	95.2
	2.00	22	4.8	4.8	100.0
Total		456	100.0	100.0	

Note that SPSS now reports 456 instead of 464 cases. The 8 cases with 3 or more missing values are still in our data but they are excluded from all analyses. We can see why in data view as shown below.

	id	mis_1	filt_1	var	var	var
19	0019	.00	1.00			
20	0020	.00	1.00			
21	0021	8.00	.00			
22	0022	.00	1.00			
23	0023	.00	1.00			

IBM SPSS Statistics Processor is ready | Unicode:ON | Filter On

- ① Case 21 has 8 missing values on q1 to q9 and we recoded this into zero on our filter variable.
- ② The strikethrough its \$casenum shows that case 21 is currently filtered out.
- ③ The status bar confirms that a filter variable is in effect. Finally, let's **deactivate** our filter by simply running

FILTER OFF.

We'll leave our filter variable filt_1 in the data. It won't bother us in any way.

Example 2 - Filter on 2 Variables

For some other analysis, we'd like to use only female respondents working in sales or marketing. A good starting point is running a very simple contingency table as shown below.

***Show values and value labels in subsequent output to set tnumbers both.**

***Show frequencies for job type per gender.**
crosstabs gender by jtype.

Result

		Current job type				
		1 Sales	2 Marketing	3 IT	4 Middle Management	5 Upper Management
gender	0 female	101	80	N = 181 RESPONDENTS ARE FEMALE AND WORK IN SALES OR MARKETING		
	1 male	71	57			
Total		172	137	92	30	12

As our table shows, we've 181 female respondents working in either sales or marketing. We'll now create a new filter variable holding only zeroes. We'll then set it to 1 for our case selection with a simple IF command.

***Create filter variable holding only zeroes.**
compute filt_2 = 0.

***Set filter to 1 for females in job types 1 and 2.**
if(gender = 0 & jtype <= 2) filt_2 = 1.

***Apply variable label.**

```
variable labels filt_2 'Filter in females working in
```

```
*Activate filter.  
filter by filt_2.
```

```
*Confirm filter working properly.  
crosstabs gender by jtype.
```

Rerunning our contingency table (not shown) confirms that SPSS now reports only 181 female cases working in marketing or sales. Also note that

we now have 2 filter variables in our data

and that's just fine but

only 1 filter variable can be active at any time.

Ok. Let's deactivate our new filter variable as well with

```
FILTER OFF.
```

Example 3 - Filter without Filter Variable

Experienced SPSS users may know that

- **TEMPORARY** can “undo” some data editing that follow it and
- **SELECT IF** permanently deletes cases from your data.

By combining them you can circumvent the need for creating a filter variable but for 1 analysis at the time only. The example below shows just

that: the first **CROSSTABS** is limited to a selection of cases but also rolls back our case deletion. The second CROSSTABS therefore includes all cases again.

```
*Make following transformation(s) temporary.  
temporary.
```

```
*Delete cases unless gender = 1 & jtype = 3.  
select if (gender = 1 & jtype = 3).
```

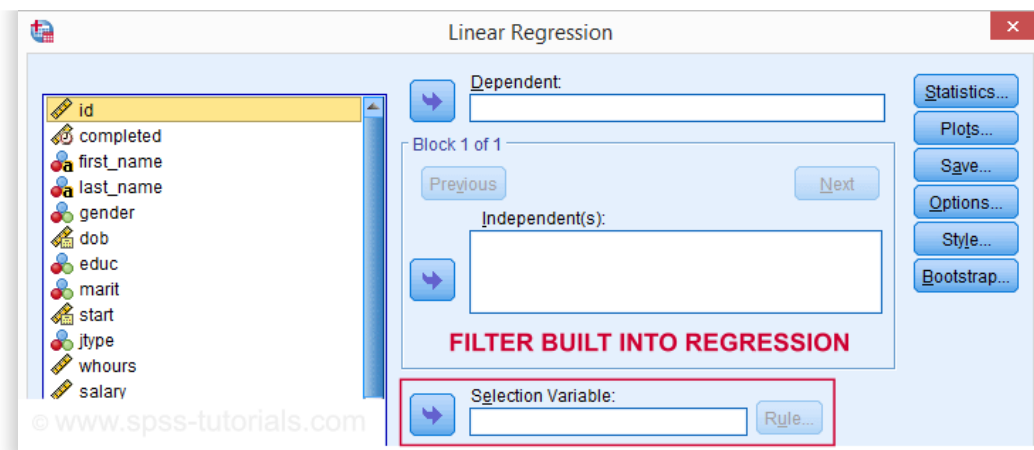
```
*Crosstabs includes only males in IT and rolls back c  
crosstabs gender by jtype.
```

```
*Crosstabs includes all cases again.  
crosstabs gender by jtype.
```

Tip - Commands with Built-In Filters

Something else you may want to know is that some commands have a built-in filter. These are

- REGRESSION,
- LOGISTIC REGRESSION,
- FACTOR and
- DISCRIMINANT.



The dialog suggests you can filter cases -for this command only- based on just 1 variable. I suspect you can enter more complex conditions on the resulting /SELECT subcommand as well. I haven't tried it.

In any case, I think these **built-in filters can be very handy** and it kinda puzzles me they're only limited to the 4 aforementioned commands.

Warning - Data Editing with Filter

Most data editing in SPSS is unaffected by filtering. For example, <https://www.spss-tutorials.com/how-to-compute-means-in-spss/> **over variables** -as shown below- affects *all* cases, regardless of whatever filter is active. We therefore need **DO IF** or **IF** to restrict this transformation to a selection of cases. However,

an active filter *does* affect functions *over* cases.

Some examples that we'll demonstrate below are

- adding a case count with **AGGREGATE**;
- computing **z-scores** for one or many variables;

- adding ranks, or with **RANK**.

SPSS Data Editing Affected by Filter Examples

***Reactivate female sales filter.**

filter by filt_2.

***Not affected by filter: add mean over q1 to q9 to data.**

compute mean_1 = mean(q1 to q9).

execute.

***Affected by filter: add case count to data.**

aggregate outfile * mode addvariables

/ofreq = n.

***Affected by filter: add z-scores salary to data..**

descriptives salary

/save.

***Affected by filter: add median groups salary to data**

rank salary

/ntiles(2) into med_salary.

Result

	filt_2	mean_1	ofreq	Zsalary	med_salary	var
13	.00	5.63
14	.00	6.56
15	1.00	6.25	181	1.09983	2	.
16	.00	4.56
17	1.00	4.89	181	.	.	.

© www.spss-tutorials.com

Right. So that's pretty much all about filtering in SPSS. I hope you found this tutorial helpful and

Thanks for reading!

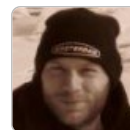
Let me know what you think!

[Done!](#)

**Required field. Your comment will show up after approval from a moderator.*

This tutorial has 16 comments

By **Ruben Geert van den Berg** on May 7th, 2019



Hi Carolyn!

Create a filter variable based on more than 1 variable. You could use something like

IF(year = 2015 & ID = 117) filt_1 =

[Expand comment](#) | [all comments](#)

1 ... 4

Ne
Va

Get In Touch!

SPSS Help (Netherlands)

Ruben Geert van den Berg

Sigma Plus Statistiek

 LinkedIn

www.sigma-plus-statistiek.nl

 Facebook

info@sigma-plus-statistiek.nl

SPSS Help (International)

SPSS tutorials

www.spss-tutorials.com

info@spss-tutorials.com

 **SPSS tutorials**

© Copyright Protected 2019 by **Sigma Plus Statistiek**

Disclaimer

