

Chapter 120

Filters

Introduction

Filters are conditional statements about data in columns that allow you to restrict analysis and graphics to certain subsets of the data. For example, suppose you have a column named Gender with two different values: "Male" and "Female". Suppose you want to perform an analysis on only the males in the dataset. You could copy and paste the male data into another dataset and run the analysis, but it would be much simpler to just impose a filter on the dataset before running the analysis.

The Filter System

NCSS offers a very powerful facility for creating, editing, and maintaining filters. The dataset can only have one filter active at a time, but a filter may have as many filter conditions involving as many columns as you like. The currently loaded filter is stored with the data file, along with whether or not it is active and whether or not filtered rows are hidden. It is easy to activate and deactivate the dataset filter anytime. Furthermore, it is easy to see which rows are filtered out because rows in the Data Table that do not pass the filter are grayed out. You can even hide filtered rows from view in the Data Table. You can load and save filters to separate NCSS filter files (*.ncssfilter) for use by various datasets.

The filter item in the Column Info Table allows you to see how the current filter is distributed across the various columns in the dataset. Columns that have a filter statement entered are indicated with a filter icon in the column heading. Columns not used in the current filter are left blank.

Types of Filter Conditions

There are three different types of conditions that can be combined to make up a filter: *List Conditions*, *Range Conditions*, and *Advanced Conditions*.

Here are some examples of the three different kinds of filter conditions:

Column	Actual Filter Condition	Column Info Table Statement	Condition Type
C2	C2=1	1	List Condition
C3	C3=1,2,10	1,2,10	List Condition
C3	C3<>5	<>5	List Condition
C4	C4>1.552	>1.552	Range Condition
C4	C4>C5	>C5	Range Condition
C4	C4<>C5	<>C5	Range Condition
C4	(C4>1) and (C4<10)	(>4)AND(<10)	Range Condition
C5	(C5<0) or (C5>15)	(<0)OR(>15)	Range Condition
NA	(C1+C2)>(C4+C5)	(C1+C2)>(C4+C5)	Advanced Condition

List Conditions

A list condition is one in which you specify a list of values to either include or exclude in analyses and plots. The list should be entered with values separated by commas.

The Filter Editor makes creating a list very simple... just select values you would like to include or exclude from the list of all possible values for the chosen filter column.

Syntax

The basic syntax of a list condition is

[Column] [Logic Operator] [Value 1], [Value 2], [Value 3], ...

With 2 possible logic operators:

- = Equal to
- <> Not equal to (≠)

If you are using a computer language setting other than English that uses a comma as the decimal symbol, then the list should be separated by the computer's specified list separator (this is usually a semicolon if not a comma). Only values, not column names, may be included in the list.

Rules:

1. Text values must be enclosed in double quotes (e.g., "Male").
2. Use the word MISSING to represent a missing value (a blank).
3. No functions allowed.

Examples:

Column	Actual Filter Condition	Column Info Table Statement
C2	C2=1	1
C3	C3=1,2,10	1,2,10
C3	C3<>5	<>5

Range Conditions

A range condition allows you to include values based on a range statement (e.g., greater than, less than, etc.). A single range statement may contain up to two range sections combined with either OR or AND. A range condition may involve other columns in the dataset.

Syntax

The basic syntax of each piece of a range condition is

[Column] [Logic Operator] [Value or Function]

With 6 possible logic operators:

- = Equal to
- <> Not equal to (≠)
- < Less than
- <= Less than or equal to (≤)
- > Greater than
- >= Greater than or equal to (≥)

Rules:

1. Combine statements using AND or OR to create bounded ranges (i.e., $10 < C1 \leq 20$ is entered as $(C1 > 10) \text{ AND } (C1 \leq 20)$).
2. Text values must be enclosed in double quotes (e.g., "Male").
3. No lists allowed if the logic operator is something other than equal to or not equal to.
4. Use the word MISSING to represent a missing value (a blank).
5. You cannot use functions that require more than one row of data such as LAG, LED, etc.
6. You cannot use special functions such as RECODE, LOCATE, etc.

Examples:

Column	Actual Filter Condition	Column Info Table Statement
C4	$C4 > 1.552$	>1.552
C4	$C4 > C5$	$>C5$
C4	$C4 < C5$	$<>C5$
C4	$(C4 > 1) \text{ and } (C4 < 10)$	$(>4) \text{ AND } (<10)$
C5	$(C5 < 0) \text{ or } (C5 > 15)$	$(<0) \text{ OR } (>15)$

Advanced Conditions

An advanced condition gives you complete control over your filter. You can enter virtually any filter statement you want. You can combine statements with AND and OR as needed. An advanced filter statement is not restricted to involve only one column but may involve multiple columns in a single statement.

If you enter an advanced filter statement in the Filter Editor, then another item will appear on the Column Info Editor. Once an advanced filter statement has been created with the Filter Editor, you can edit it directly on the Column Info Editor.

The syntax for advanced conditions is the same as that for range conditions.

Examples:

Column	Actual Filter Condition	Column Info Table Statement
NA	$(C1+C2)>(C4+C5)$	$(C1+C2)>(C4+C5)$
NA	$((C1>C2)OR(C3="Male"))AND(C4>100)$	$((C1>C2)OR(C3="Male"))AND(C4>100)$

Combining Filter Conditions

If your data filter has more than one condition, then you must specify how to combine these condition statements in a single filter. That is, you must decide whether to keep a row only if ALL conditions are True or if ANY of the statements is True. This is the same as combining the conditions together with AND and OR, respectively. You are allowed to combine filter conditions in one of the two ways, but not at the same time.

On the Filter Editor window this option is presented as *Keep a dataset row if:* with choices:

- ALL of the following conditions are true**
 Filter conditions will be combined with "AND" to create the final filter statement. This option is the strictest and will filter out more rows.
- ANY of the following conditions is true (at least one)**
 Filter conditions will be combined with "OR" to create the final filter statement. A row will pass the filter if any of the conditions is true. This option is less strict and will filter out fewer rows.

Editing the Data Filter

It is easy to create and edit filters using the Filter Editor window. The Filter Editor is a simple point-and-click tool for creating and editing filters. Remember that you do not have to use the filter editor to create a filter; you can type any filter directly into the Column Info Table.

There are several ways to launch the filter editor.

1. From the Data Window (or Procedure Window) main toolbar, click on the *Filter* dropdown button and select *Load the Filter Editor*.
2. From the Data Window menu, select *Data > Filter > Edit Filter*.
3. From the Data Window menu, select *Data > Column X Info > Filter* to create/edit a filter condition for the column containing the active cell.
4. Click on the button at the right side of a filter cell when it is selected in the Column Info Table to create/edit a filter condition for the column containing the active cell.
5. Select a cell in the Data Table, right-click, and select *Column X Info > Edit Filter* from the popup menu.

Activating and Deactivating Filters

The *Filter Active* option in the Filter button dropdown on the Data Window (or Procedure Window) main toolbar controls whether the currently loaded filter is active. Any rows for which the filter statement does not evaluate to True are excluded from analyses and plots. Filtered rows are grayed out in the Data Table. Rows that pass the filter are indicated with a bold, green number in the row heading. Filtered values are not deleted from the dataset. You can activate and deactivate the filter in between analyses as often as you would like.

Hiding and Showing Filtered Rows

The *Hide Filtered Rows* in the Filter button dropdown on the Data Window (or Procedure Window) main toolbar determines whether filtered rows will be hidden from view. When the filter is inactive, all rows are always shown. When filtered rows are hidden, they are not deleted from the dataset--just hidden from view. You can see which rows are filtered out by unchecking the option. Filtered rows will be grayed out, but visible when this option is unchecked.