Chapter 103

The Data Window

Introduction

This chapter discusses the operation of the **NCSS** *Data Window*, one of the four main windows of the **NCSS** statistical analysis system. The other three windows are the *Output Window*, the *Procedure Window*, and the *Procedure Navigator Window*. These are described in other chapters. The Data Window is the first screen that you are presented with when you start the program. The program terminates when the Data Window is closed. The Data Window lets you enter, view, and modify the information contained in an **NCSS** *Dataset*. Datasets are stored as data files with the extension *.NCSS* (e.g. "SampleData.NCSS").

The Data Window is comprised of two main data entry sections: the *Data Table* and the *Column Info Table*. The Column Info Table has a yellow background and can be hidden, while the Data Table has a white background and is always visible. Both tables have the familiar "spreadsheet" appearance. In fact, the operation of the Data and Column Info tables is similar to the operation of other spreadsheets with which you are probably already familiar, like Microsoft Excel. You can type values in each cell, cut, copy, and paste data, etc. just like in Microsoft Excel. The main difference, however, between the **NCSS** Data Table and a spreadsheet like Excel is that the **NCSS** Data Table is column-based, while a spreadsheet is cell-based. In an Excel spreadsheet you can enter formulas into a single cell and calculate results based on a range of other cell, either in a column or in a row. The **NCSS** Data Table allows formulas that apply only to an entire column, called transformations. When changing the display font of the **NCSS** Data Table, the font will apply to an entire column, not just a range of cells in the column. You can, however, copy and paste individual cells or ranges, just like you can in a spreadsheet.

The Data Window also contains a *Summary Statistics Table* with a light-blue background that automatically calculates user-selected statistics for each column. The statistics are updated each time a change is made to the data in a column in the Data Table. Click *Statistics & Formatting* on the right side of the Summary Statistics Table title bar to select from over 40 available summary statistics and to adjust the display formatting.

In addition to the Data, Column Info, and Summary Statistics tables, the Data Window has various other components that assist the user in modifying, viewing, editing, and saving data. These are the *Menu*, the *Main Toolbar*, the T*ransformation Toolbar*, the *Bottom Bar*, and the *Data Entry Bar*. Each of these components will be described in the sections that follow.

Datasets

NCSS analyzes data contained in an **NCSS** *Dataset*. An **NCSS** dataset contains all of the information in the Data Table (including column and row widths), the Column Info Table, the dataset notes, the currently-loaded filter, the filter active option, the hide filtered rows option, and the named value-label lists associated with the data. Datasets are stored as *Data Files* and have the extension *.NCSS* (e.g. "SampleData.NCSS"). When the program asks for a *variable* in an analysis or graphics routine, it is actually asking you to choose a column from the dataset.

NCSS accepts both numeric and text data. Numeric data may contain up to 16 digits (double-precision). Text data may contain up to 32000 characters per cell.

Data Limits

An **NCSS** dataset is limited to 2 billion rows and to 2 billion columns; this essentially means that a dataset is limited only by the available memory on your computer. You will likely run out of memory long before the dataset reaches these limits.

By default, the dataset starts with 256 columns and 65536 rows available. You can change the default number of rows and columns by selecting *File > System Options* from the Data Window menu and choosing the *Data* tab on the System Options window. You can add or delete rows and columns from the dataset at any time.

Adding and Deleting Rows and Columns

When running data operations like import, fill, paste, etc., **NCSS** will determine the number of rows and columns required to complete the operation and automatically add more rows and columns if necessary. To manually add, insert, and delete rows or columns from the Data Table at any time, click *Edit* on the Data Window menu, and then select the appropriate option from the drop-down menu.

Accuracy

NCSS maintains double-precision (sixteen-digit) accuracy in the data values. When you apply a special format to a column, only the display of the data is modified. The actual sixteen-digit number is maintained.

All numbers are stored in the IEEE floating-point format. Such numbers range from 4.94E-324 to 1.798E308 for positive values, 0, and from -1.798E308 to -4.940E-324 for negative values.

Because of the rounding that has to occur to fit into the IEEE format, it is impossible to express all numbers exactly. For example, the common fraction of one-third cannot be expressed exactly, but must be rounded at the last digit. This rounding may cause strange results for certain decimal numbers. For example, you might enter 453.4537 and have it displayed as 453.45369999999999. There is nothing we can do to change this oddity. You can rescale your data by multiplying the variable by 10 or 100. However, this problem rarely occurs in practice and will not change the accuracy of your results.

The "E" notation is used for expressing large and small numbers in scientific notation. The number after the E is the exponent (base 10) that is applied to the base number. Thus, 3.238E-03 means 0.003238.

Other examples are:

E-Notation	Decimal Equivalent
-3.238E-06	-0.000003238
-3.238E-02	-0.03238
3.238E04	32380.0
3.238E06	3238000.0

Missing Values

Missing values are represented by empty cells. In procedures that use only numeric values, such as the average, text values are also treated as missing values. Unless otherwise specified, **NCSS** uses *row-wise deletion* of observations with missing values. This means that if any of the active variables in a row has a missing value, the whole observation is omitted from the calculations.

The Data Table

The Data Table contains all of the data values in your dataset that will be analyzed and included in plots. The way values are displayed and interpreted on the Data Table is controlled by the information in the Column Info Table, specifically the column data type and format. The Data Table is column-based; data types, formats, formulas, fonts, etc. apply to entire columns not individual cells.

Sorting Data

To sort data in the Data Table, select *Data > Sort* from the Data Window menu. Alternatively, you can click on the *Sort* button on the main toolbar. This loads a dialog that allows you to select up to three columns to sort on. The entire database is sorted when you perform this operation, not just the columns chosen to sort by. The contents of each individual row will remain the same; the rows will just be rearranged according to the sort columns.

Filling Columns with New Data

You can fill columns with data using the *Fill* window. To open the Fill window, select *Data > Fill* from the Data Window menu. Alternatively, you can click on the *Fill* button on the main toolbar. This loads a dialog that allows you to fill one or more columns with a single value, a numeric series, a repeating sequence, or uniform and normal random numbers. You can fill as many rows as you would like. If you choose to fill a number of rows that is greater than the number of the last row in the dataset, **NCSS** will automatically add additional rows as necessary to complete the fill operation.

Resizing Rows and Columns

You can resize and row or column by placing your mouse pointer on the right edge of a column heading or the bottom edge of a row heading. Left-click and drag your mouse to increase or decrease the height (width) of a row (column). If multiple rows or columns are selected when you resize, they will all be resized equally. You can also automatically resize all columns using the resizing options available in the Data Window *Edit* menu.

Column Font

Change the font of any column by clicking on *Data > Column X Font*. This will load a window to change the font of the currently-active column in the Data Table. This font change only affects display in the Data Table, not the output. The entire column will be affected. You cannot change the font for individual cells, only for entire columns. The font does not affect the display of values in the Column Info and Summary Statistics Tables.

Zoom

To increase or decrease the size of all text in the Data Table, the Column Info Table, and the Summary Statistics Table, select *View > Zoom* from the Data Window menu. Alternatively, you can use the zooming tools available in the lower right side of the screen on the bottom bar. If the bottom bar is not showing, select *View > Bottom Bar* to make it appear. You can adjust the table zoom to any value between 40% and 400%.

The Column Info Table

The Column Info Table contains information about each column in the dataset: *Name*, *Label*, *Data Type*, *Format*, *Value Labels*, *Value Order*, *Transformation*, *Note*, and *Filter*. These fields control how the data will be displayed in the Data Table and how the data will be interpreted by analysis and graphics procedures. To view or hide the Column Info Table Editor, click on the blue *Column Info Tab* in the upper-left corner of the screen or select *View > Column Info Editor* from the menu.

The Column Info Table is displayed using either one of two possible views… the long view or the short view. The long view displays information about each column row-by-row; a row in the long-view Column Info Table corresponds to a column in the Data Table. The short view is a column-by-column listing of the information about each data column in the Data Table. The short view has the advantage that you can always see the data below the column information. The long view has the advantage of wider cells for each column information object. You should use the view that is most comfortable for you. You can easily switch between views by clicking on the *Rotate View* button above the Column Info Table or by selecting *View > Rotate Column Info Editor View* from the menu.

To change column information for any given column, directly edit each cell in the Column Info Table or use the editor windows that are available for each column information object. You can load an editor for each cell by clicking on the button that will appear on the right side of the cell rectangle when focused on any cell in the table. The editor will guide you through data entry in a point-and-click manner, making it simple to make changes. We suggest that you begin by using the tools for each column information object and as you get more comfortable with the syntax and contents of each cell, start typing in information directly into the table.

Number

The first item in the Column Info Table is the *number* of each data column. In the short view, the number is displayed at the top. In the long view, the number is displayed at the far left. You cannot edit this value directly. Although you usually refer to a data column by its name when you write transformations and make variable selections in procedures, your selection may be referenced in analysis procedures by number. This is all handled internally, so you will not have to worry about it except when you move columns around within a dataset. When you do this, you will have to check stored procedure templates to make sure they still refer to the correct columns.

Name

The next item in the Column Info Table is the *name* column. Each column is given a default name when the dataset is created. Throughout the program, you refer to the column by its name. The default names are C1, C2, C3, etc. Each column must have a unique name.

A column's name may be changed at any time by editing the Name item in the Column Info Table. The syntax for the naming of columns follows these rules:

1. Column names must begin with a letter.

2. Column names can contain only letters, numbers, or an underscore.

3. Column names may not include spaces.

4. Column names may not include mathematical symbols.

5. Upper- and lower-case letters are treated the same. Names are not case-sensitive.

Label

The next item in the Column Info Table is the *Label* column. By default, the column labels are left empty. Column labels can be as long as you want and contain any characters you like, including spaces. No attempt is made to trim this label if it is too long for the space provided to display it in a particular report. Each procedure has an option in which you designate whether or not to include these labels on the output.

Data Type

The next item in the Column Info Table is the *Data Type* column. There are three possible data types: *General (Numeric or Text)*, *Text*, and *DateTime*. The data type determines how data in the Data Table will be stored, interpreted, and displayed in output reports and plots, particularly when the column is used as a grouping or break variable in a procedure such as cross-tabulation or descriptive statistics. The Data Type is ignored when the column is used in numeric calculations, i.e. numbers stored in a column designated with a Text or DateTime data type will still be interpreted as numbers.

When you change a column's data type, any data values in the Data Table will be converted to the new data type, so be careful. When converting from the text data type to General or DateTime, some information may be lost when long text numbers are converted to double-precision numbers or date values.

We will now describe each data type.

General (Numeric or Text)

The *General* data type is the default data type for **NCSS** data. Data may either be text values or numeric. Dates are treated as text. The actual appearance of the data values on the spreadsheet depends on the format statement. By default, numbers will be displayed with up to 11 digits, depending on the column widths. Regardless of the data type and display format, the actual sixteen-digit numbers are always maintained with the dataset.

Text

The *Text* data type is useful when you have large numeric identification values that you do not want converted to 16-digit floating-point numbers. Numeric text values will still be interpreted as numbers in transformations and calculations for data analysis. There is no special formatting for text values.

For example, the value 9876543212345678987654321 entered into a column with data type = General will be converted to the maximum floating-point representation of 9.87654321234568E+24. Some of the digits are removed because a floating point number can only store up to 16 digits. The same value entered into a column with data type = Text will maintain the full value 9876543212345678987654321 because it is stored as a text value.

DateTime

The *DateTime* data type is useful when you want to format numeric values as dates. DateTime values are considered numeric in transformation and data analysis calculations. Data values in DateTime data columns are still stored with 16-digit floating point accuracy. By default, dates are formatted as a "short date", which depends on your computer's language setting.

Format

The next item in the Column Info Table is the *Format* column. The application of the format depends on the column's data type. The column format only affects how a value is displayed in the Data Table, not in reports and plots. It does not affect calculations. For example, suppose the value 4.443566 is stored in column 1. If I change the display format for column 1 to "0.00" then the value will be displayed as 4.44 on the Data Table. However, when this value is used in calculations, the entire number 4.443566 will still be used.

The simplest way to edit a column's display format is by loading the format editor. To launch the editor, click on the button at the right side of a format cell when it is selected in the Column Info Table or from the Data Window menu, select *Data > Column X Info > Format*.

 This requires no knowledge of format syntax. Several built-in formats are available and it is easy to write your own custom format. Each custom format can have as many as four sections, separated by semicolons:

1. One for positive numbers

2. One for negative numbers

3. One for zeros

4. One for text

Each section is optional, but if you have more than one, you define the format section by the placement of extra semicolons. If you only use one format section, it defines the format for all numbers, both positive and negative.

Following is an example of a typical, four-section format:

0.000;(0.000);0;[Red]

We will now describe the format statements for each data type.

General (Numeric or Text)

Non-numeric text values are not formatted. Numeric values are formatted as described in the following table.

Table of Numeric Formatting Statements

Statement	Name	Description
C or **c**	Currency	Number is formatted as a currency amount. The formatted value depends on the user's current computer language settings.
E or **e**	Scientific (exponential)	The number is formatted as "-d.ddd…E+ddd" or "-d.ddd…e+ddd", where each 'd' indicates a digit (0-9). The result starts with a minus sign if the number is negative. One digit always precedes the decimal point. Precision specifier (e.g. "E5") allowed. The case of the format specifier indicates whether to prefix the exponent with an 'E' or an 'e'. The exponent always consists of a plus or minus sign and a minimum of three digits. The exponent is padded with zeros to meet this minimum, if required.
F or **f**	Fixed-Point	The number is formatted as "-ddd.ddd…" where each 'd' indicates a digit (0-9). The formatted value starts with a minus sign if the number is negative. Precision specifier (e.g. "f4") allowed.
G or **g**	General	The number is formatted as the most compact of either fixed-point or scientific notation, depending on the type of the number and whether a precision specifier is present. Precision specifier (e.g. "G5") allowed.
N or **n**	Number	The number is formatted as "-d,ddd,ddd.ddd…", where '-' indicates a negative number symbol if required, 'd' indicates a digit (0-9), ',' indicates a thousand separator between number groups, and '.' indicates a decimal point symbol. The actual negative number pattern, number group size, thousand separator, and decimal separator are specified by the user's current computer language settings. Precision specifier (e.g. "N5") allowed.
P or **p**	Percent	The number is formatted as a percent as defined by the user's current computer language settings. The converted number is multiplied by 100 in order to be presented as a percentage. Precision specifier (e.g. "{p6}") allowed.

Table of Numeric Formatting Statements (continued)

Statement	Name	Description
0	Zero Placeholder	If the value being formatted has a digit in the position where the '0' appears in the format statement, then that digit is copied to the result. The position of the leftmost '0' before the decimal point and the rightmost '0' after the decimal point determines the range of digits that are always present in the result. The "00" specifier causes the value to be rounded to the nearest digit preceding the decimal, where rounding away from zero is always used. For example, formatting 34.5 with "00" would result in the value 35.
#	Digit Placeholder	If the value being formatted has a digit in the position where the '#' appears in the format statement, then that digit is copied to the result. Otherwise, nothing is stored in that position in the result. This specifier never displays the '0' character if it is not a significant digit, even if '0' is the only digit in the statement. It will display the '0' character if it is a significant digit in the number being displayed. The "##" format statement causes the value to be rounded to the nearest digit preceding the decimal, where rounding away from zero is always used. For example, formatting 34.5 with "##" would result in the value 35.
.	Decimal Point	The first '.' character in the format statement determines the location of the decimal separator in the formatted value; any additional '.' characters are ignored. The actual character used as the decimal separator is determined by the user's current computer language settings that control formatting.
,	Thousand separator and number scaling	**Thousand Separator Specifier** If one or more ',' characters is specified between two digit placeholders (0 or #) that format the integral digits of a number, a group separator character is inserted between each number group in the integral part of the output. The user's current computer language settings that control formatting determine the character used as the number group separator and the size of each number group. **Number Scaling Specifier** If one or more ',' characters is specified immediately to the left of the explicit or implicit decimal point, the number to be formatted is divided by 1000 each time a number scaling specifier occurs. For example, if the statement "0,," is used to format the number 100 million, the output is "100". You can use thousand separator and number scaling specifiers in the same format statement.
%	Percentage Placeholder	The presence of a '%' character in a format statement causes a number to be multiplied by 100 before it is formatted. The appropriate symbol is inserted in the number itself at the location where the '%' appears in the format statement. The percent character used is dependent on the user's current computer language settings that control formatting.
E0 **E+0** **E-0** **e0** **e+0** **e-0**	Scientific notation	If any of the values "E", "E+", "E-", "e", "e+", or "e-" are present in the format statement and are followed immediately by at least one '0' character, then the number is formatted using scientific notation with an 'E' or 'e' inserted between the number and the exponent. The number of '0' characters following the scientific notation indicator determines the minimum number of digits to output for the exponent. The "E+" and "e+" formats indicate that a sign character (plus or minus) should always precede the exponent. The "E", "E-", "e", or "e-" formats indicate that a sign character should only precede negative exponents.
'ABC' **"ABC"**	Literal	Characters enclosed in single or double quotes are copied to the result, and do not affect formatting.
;	Section separator	The ';' character is used to separate sections for positive, negative, zero numbers, and text in the format statement. The four sections of a custom format are for positive numbers, negative numbers, zeros, and text.
[color]	Color	Displays the value in the specified color.
[conditional value]	Conditional Formatting Statement	Under normal circumstances, the four sections of a custom format are for positive numbers, negative numbers, zeros, and text. Use brackets to indicate a different condition for each section. For example, you might want numbers less than three to be black and those greater than three to be red. You would use: [>3] [Red][General];[<-3][Red][General];[Black][General]

Numeric Formatting Examples

Statement	Value	Formatted Result
C	12345.6789	$12,345.68
C0	12345.6789	$12,346
e2	12345.6789	1.23e+004
F	12345.6789	12345.68
F6	12345.6789	12345.678900
G	12345.6789	12345.6789
G7	12345.6789	12345.68
G	0.0000023	2.3E-06
G2	1234	1.2E+03
N	12345.6789	12,345.68
N5	12345.6789	12,345.67890
P	0.126	12.60 %
#####	123	123
00000	123	00123
(###) ### - ####	1234567890	(123) 456 - 7890
#.##	1.2	1.2
0.00	1.2	1.20
#,#	1234567890	1,234,567,890
#,,	1234567890	1235
#0.##%	0.086	8.6%
0.###E+0	86000	8.6E+4
##;(##)	1234 -1234	1234 (1234)
[Red];[Blue];[Green]0.00;[Black]	123 -123 0 Hello	123 -123 0.00 Hello
[>10][Red];[<10][Blue];[=10][Green];[Text][Purple]	20 5 10 Hello	20 5 10 Hello

Text

Text values are not formatted.

DateTime

Non-numeric text values are not formatted, but are displayed as entered. Numeric values are formatted as dates and/or times as described in the following table.

Table of Date/Time Formatting Statements

Statement	Name	Description
d	Short Date	The date is formatted as a short date. The formatted value depends on the user's current computer language settings.
D	Long Date	The date is formatted as a long date. The formatted value depends on the user's current computer language settings.
f	Full Date/Time (Short Time)	Represents a combination of the long date (D) and short time (t) patterns, separated by a space. The formatted value depends on the user's current computer language settings.
F	Full Date/Time (Long Time)	Represents a combination of the long date (D) and long time (t) patterns, separated by a space. The formatted value depends on the user's current computer language settings.
g	General Date/Time (Short Time)	Represents a combination of the short date (d) and short time (t) patterns, separated by a space. The formatted value depends on the user's current computer language settings.
G	General Date/Time (Long Time)	Represents a combination of the short date (d) and long time (T) patterns, separated by a space. The formatted value depends on the user's current computer language settings.
M or **m**	Month/Day	The date is formatted as the month followed by the day, separated by a space. The formatted value depends on the user's current computer language settings.
t	Short Time	The date is formatted as a short time. The formatted value depends on the user's current computer language settings.
T	Long Time	The date is formatted as a long time. The formatted value depends on the user's current computer language settings.
Y or **y**	Year/Month	The date is formatted as the month followed by the year, separated by a space. The formatted value depends on the user's current computer language settings.
d	Custom	Represents the day of the month as a number from 1 through 31. A single-digit day is formatted without a leading zero. (Example: M/d/yyyy)
dd	Custom	Represents the day of the month as a number from 01 through 31. A single-digit day is formatted with a leading zero. (Example: MM/dd/yyyy)
ddd	Custom	Represents the abbreviated name of the day of the week. The formatted value depends on the user's current computer language settings. (Example: ddd, MMM d, yyyy)
dddd	Custom	Represents the full name of the day of the week. The formatted value depends on the user's current computer language settings. (Example: dddd, MMMM d, yyyy)
M	Custom	Represents the month as a number from 1 through 12. A single-digit month is formatted without a leading zero. <u>Must be capitalized</u>. (Example: M/d/yyyy)
MM	Custom	Represents the month as a number from 01 through 12. A single-digit month is formatted with a leading zero. <u>Must be capitalized</u>. (Example: MM/dd/yyyy)
MMM	Custom	Represents the abbreviated name of the month. The formatted value depends on the user's current computer language settings. <u>Must be capitalized</u>. (Example: ddd, MMM d, yyyy)
MMMM	Custom	Represents the full name of the month. The formatted value depends on the user's current computer language settings. <u>Must be capitalized</u>. (Example: dddd, MMMM d, yyyy)
y	Custom	Represents the year as at most a two-digit number. If the year has more than two digits, only the two low-order digits appear in the result. If the year has fewer than two digits, the number is formatted without a leading zero. (Example: M/d/y)
yy	Custom	Represents the year as a two-digit number. If the year has more than two digits, only the two low-order digits appear in the result. If the year has fewer than two digits, the number is padded with leading zeroes to achieve two digits. (Example: MM/dd/yy)

Table of Date/Time Formatting Statements (continued)

Statement	Name	Description
yyy	Custom	Represents the year as a three-digit number. If the year has more than three digits, only the three low-order digits appear in the result. If the year has fewer than three digits, the number is padded with leading zeroes to achieve three digits. (Example: ddd, MMM d, yyy)
yyyy	Custom	Represents the year as a four-digit number. If the year has more than four digits, only the four low-order digits appear in the result. If the year has fewer than four digits, the number is padded with leading zeroes to achieve four digits.(Example: dddd, MMMM d, yyyy)
yyyyy	Custom	Represents the year as a five-digit number. If the year has more than five digits, only the five low-order digits appear in the result. If the year has fewer than five digits, the number is padded with leading zeroes to achieve five digits. If there are additional "y" specifiers, the number is padded with as many leading zeroes as necessary to achieve the number of "y" specifiers. (Example: dddd, MMMM d, yyyyy)
h	Custom	Represents the hour as a number from 1 through 12, that is, the hour as represented by a 12-hour clock that counts the whole hours since midnight or noon. A single-digit hour is formatted without a leading zero. (Example: h:m)
hh	Custom	Represents the hour as a number from 01 through 12, that is, the hour as represented by a 12-hour clock that counts the whole hours since midnight or noon. A single-digit hour is formatted with a leading zero. (Example: hh:mm)
H	Custom	Represents the hour as a number from 0 through 23, that is, the hour as represented by a zero-based 24-hour clock that counts the hours since midnight. A single-digit hour is formatted without a leading zero. (Example: H:m)
HH	Custom	Represents the hour as a number from 00 through 23, that is, the hour as represented by a zero-based 24-hour clock that counts the hours since midnight. A single-digit hour is formatted with a leading zero. (Example: HH:mm)
m	Custom	Represents the minute as number from 0 through 59. The minute represents whole minutes passed since the last hour. A single-digit minute is formatted without a leading zero. (Example: h:m)
mm	Custom	Represents the minute as number from 00 through 59. The minute represents whole minutes passed since the last hour. A single-digit minute is formatted with a leading zero. (Example: hh:mm)
s	Custom	Represents the seconds as a number from 0 through 59. The second represents whole seconds passed since the last minute. A single-digit second is formatted without a leading zero. (Example: h:m:s)
ss	Custom	Represents the seconds as a number from 00 through 59. The second represents whole seconds passed since the last minute. A single-digit second is formatted with a leading zero. (Example: hh:mm:ss)
t	Custom	Represents the first character of the A.M./P.M. designator. The formatted value depends on the user's current computer language settings. (Example: h:m:s t)
tt	Custom	Represents the A.M./P.M. designator. The formatted value depends on the user's current computer language settings. (Example: hh:mm:ss tt)
%c	Single Custom Specifier	Represents the result associated with a custom format specifier "c", when the custom format statement consists solely of that custom format specifier. (Example: %y)
\c	Escape Character	The escape character. Displays the character "c" as a literal when that character is preceded by the escape character (\). To insert the backslash character itself in the result string, use two escape characters ("\\").
:	Time Separator	Represents the time separator that is used to differentiate hours, minutes, and seconds. The formatted value depends on the user's current computer language settings.
/	Date Separator	Represents the date separator that is used to differentiate days, months, and years. The formatted value depends on the user's current computer language settings.
'ABC' **"ABC"**	Literal	Characters enclosed in single or double quotes are copied to the result, and do not affect formatting.
;	Section separator	The ';' character is used to separate sections for positive, negative, zero numbers, and text in the format statement. The four sections of a custom format are for positive numbers, negative numbers, zeros, and text.
[color]	Color	Displays the value in the specified color. Use "[Text]" to indicate colors for text values.
[conditional value]	Conditional Formatting Statement	Use brackets to indicate a different condition for each section. (Example: [>3/15/12][Red];[<=3/15/12][Blue];[Text][Green])

Date/Time Formatting Examples

Statement	Value	Formatted Result
d	March 19, 2012 9:18:12 AM	3/19/2012
D	March 19, 2012 9:18:12 AM	Monday, March 19, 2012
f	March 19, 2012 9:18:12 AM	Monday, March 19, 2012 9:18 AM
F	March 19, 2012 9:18:12 AM	Monday, March 19, 2012 9:18:12 AM
g	March 19, 2012 9:18:12 AM	3/19/2012 9:18 AM
G	March 19, 2012 9:18:12 AM	3/19/2012 9:18:12 AM
M	March 19, 2012 9:18:12 AM	March 19
t	March 19, 2012 9:18:12 AM	9:18 AM
T	March 19, 2012 9:18:12 AM	9:18:12 AM
Y	March 19, 2012 9:18:12 AM	March, 2012
M/d/yyyy	March 19, 2012 9:18:12 AM	3/19/12
MM/dd/yy	March 19, 2012 9:18:12 AM	03/19/12
ddd, MMM d, yyy	March 19, 2012 9:18:12 AM	Mon, Mar 19, 2012
dddd, MMMM d, yyyy	March 19, 2012 9:18:12 AM	Monday, March 19, 2012
dddd, MMMM d, yyyyy	March 19, 2012 9:18:12 AM	Monday, March 19, 02012
h:m:s	March 19, 2012 9:18:12 AM	9:18:12
H:m:s	March 19, 2012 9:18:12 PM	21:18:12
hh:mm:ss t	March 19, 2012 9:18:12 AM	09:18:12 A
HH:mm:ss tt	March 19, 2012 9:18:12 PM	21:18:12 PM
%M	March 19, 2012 9:18:12 AM	3
%d	March 19, 2012 9:18:12 AM	19
M/d/yyyy[Red];[Text][Blue]	March 19, 2012 Hello	3/19/2012 Hello
[>3/15/12]M/d/y[Red];[<=3/15/12]D[Blue]; [Text][Green]	3/20/2012 3/15/2012 Hello	3/20/12 Thursday, March 15, 2012 Hello

Value Labels

The next item in the Column Info Table is the *Value Labels* column. Value labels are labels that are displayed in the place of the original value in reports and plots. For example, you might be tabulating a survey in which a "Yes" was coded as a 1 and a "No" was coded as a 0. The final report will be much more interpretable if "Yes" and "No" are displayed rather than 1 and 0. One approach would be to generate a new variable, replacing the 1's and 0's with Yes's and No's. A simpler approach, however, is to use Value Labels.

Creating Value Labels

Value labels are easy to create using the Value Label Editor tool. To launch the tool, click on the button at the right side of a value label cell when it is selected in the Column Info Table or from the Data Window menu, select *Data > Column X Info > Value Labels*. There are three ways to specify value labels for a column: *List*, *Named List*, and *File*. We will now explain each method.

List

The list is the simplest method for inputting value labels. Use this option to enter value labels that are unique to a single column in the dataset. In the table on the Value Label Editor window, simply enter values in the first column and corresponding labels in the second column. The Value Label Editor will return a properly-formatted value label list to the Column Info Table.

You can enter value labels directly on the Column Info Table using the syntax:

[*Value 1*]=[*Label 1*], [*Value 2*]=[*Label 2*], [*Value 3*]=[*Label 3*], …

where [*Value i*] and [*Label i*] represent actual values and labels, respectively. 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). If a value or label includes a space, you should enclose the entire value or label in double quotes.

Here is an example of a properly-formatted value label list as would be entered on the Column Info Table:

0 = No, 1 = Yes, 2 = "Maybe So"

Named List

A named list is useful when you have more than one column which will use the same value label list. The named list is the same as a list as described above, except that it is referenced on the Column Info Table by its name, not its content. This allows you to use the same list for several columns. If you change the values or labels on the named list, the changes will apply to all columns that use it. Named lists are stored as part of the dataset.

You create a named list using the Value Label Editor exactly the same way you create a simple list. In the table on the tool window, simply enter values in the first column and corresponding labels in the second column. The Value Label Editor will return to the Column Info Table an appropriate reference to the named list.

You can enter a named list reference directly on the Column Info Table using the syntax:

List([*List Name*])

where [*List Name*] represents the actual list name.

Here is an example of a properly-formatted value label named-list reference as would be entered on the Column Info Table:

List(YesNo)

File

Value label files are useful when you have value labels stored in separate text files to be used with multiple **NCSS** data files. Values label files are easily constructed using Windows Notepad or similar text editor. The files contain a list of values and their corresponding labels, one set per row. The value and the label must be separated by a tab (not a blank since the labels may contain blanks). Note that this is the format of data that is copied and pasted into a text file from a spreadsheet such as **Excel** or **NCSS**.

Here is an example of a properly-formatted value label file called Likert.txt:

1   Strongly Disagree
2   Disagree
3   Neutral
4   Agree
5   Strongly Agree

To label values in a dataset column using a value label file, select *File* on the Value Label Editor window, and then click *Browse* to select a value label file located on your computer or network. The file name and path will be displayed in the *Value Label File Name* box. You must then choose how to reference the file on the Column Info Table.

The options are:

- **Reference Only the File Name in Column Info**

  The value label file must be located in the same folder as the current **NCSS** data file. You would select this option if you want to have everything related to this dataset in a single folder that you might copy to another computer or send to another user.

- **Reference the Full Path and File Name in Column Info**

  The value label file can be located anywhere. You would select this option if you want to use the same value label file for several **NCSS** data files that may not reside in the same folder.

The Value Label Editor will return to the Column Info Table an appropriate reference to the value label file.

You can enter a named list reference directly on the Column Info Table using the syntax:

File([*File Name*])

where [*File Name*] represents the file name or file name and path.

Here are examples of properly-formatted value label file references as would be entered on the Column Info Table:

```
File(YesNo.txt)
File(C:\NCSS\Data\YesNo.txt)
```

Displaying Value Labels on the Data Table

To display the value labels loaded for dataset directly on the Data Table, select *View > Display Value Labels in the Data Table* from the menu. The associated label for each value will be displayed. The underlying data values are not changed. When you copy cells that display value labels, the actual values not the labels will be copied.

Using Value Labels in Reports and Plots

Even if you have entered value labels for a particular column, they will not automatically be used in reports and plots unless you make the choice to use them. On each procedure window the option to use or not to use value labels is located on the Reports tab under *Report Options*. To use value labels on a report or plot, choose "Value Labels" or "Both" for the *Value Labels* option. All values for which an appropriate label is specified will be replaced with the value label. Any values for which labels are not specified will not be affected.

Value Order

The next item in the Column Info Table is the *Value Order* column. This item allows you to specify the order in which values will be output in reports. It also allows you to specify the order in which group colors, lines, and symbols will be applied in plots. For example, suppose I have a column with values "Low" and "High". By default, "High" will be assigned to group 1 and "Low" will be assigned to group 2, because "High" comes before "Low" in alphabetical order. However, I might want "Low" to appear before "High" in reports and plots when using this column because this would represent a natural ordering based on magnitude. To do this I would enter a custom value order for this column.

You can easily create a custom value order using the Value Order Editor tool. To launch the tool, click on the button at the right side of a value order cell when it is selected in the Column Info Table or from the Data Window menu, select *Data > Column X Info > Value Order*.

The Value Order Editor gives you two options to determine how values will be ordered:

- **Alphanumeric Order (default)**

    This is the default ordering method for all columns. Values are sorted alphanumerically in reports and plots.

- **Custom Order**

    Select this option to create a custom value order. Values in reports and plots will be arranged in the order specified in the box. Click on the arrows at the right to move values up and down in the ordered list. Any values in the column that are not included in the list will be sorted alphanumerically after all values in the list have been placed in the specified order.

You can enter value orders directly on the Column Info Table using the syntax:

<div align="center">

[*Value 1*], [*Value 2*], [*Value 3*], …

</div>

where [*Value i*] represents actual values. We suggest you put values with spaces inside double quotes, but it is not required. 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).

Here is an example of a properly-formatted value order list as would be entered on the Column Info Table:

No, Yes, "Maybe So"

Transformation

The next item in the Column Info Table is the *Transformation* column. For more information about transformations and specific function syntax, see the *Transformations* chapter.

Transformations generate new data. They can be used to perform mathematical operations such as addition, multiplication, and exponentiation on columns. For example, you might want to create a new column that is the logarithm of the data in another column, the total of three columns, or the recoding of another column. A rich set of text transformations is also available.

A transformation formula specifies how the new column is constructed from the existing columns. The formula is made up of mathematical operators, variables, constants, and common mathematical functions. It is important to understand that these transformations modify columns of data, not individual cells. This is one place where the **NCSS** Data Table is different from traditional spreadsheets.

There are three general types of data transformations available: *Transformations (General)*, *Conditional (If-Then) Transformations*, and *Recode Transformations*. You can type any transformation formula directly into the Column Info Table, or you can enter it using one of three different Transformation Editor windows.

Here are some examples of the three different kinds of transformations:

Result Column	Transformation Formula	Type
C2	C1*100	General
C2	log(C1)	General
C3	C1+C2	General
C3	Sqrt(C1/C2)	General
C4	IfThen(C3=1, C1+C2; Else, 0)	Conditional (If-Then)
C4	Recode(C3; 1="Male", 2="Female")	Recode

We will now present a broad overview of transformations in **NCSS**. For more information about transformations and specific function syntax, see the *Transformations* chapter.

Transformation Components

All transformations are made up of two components: the *result column* and the *transformation formula* or *function*. Each transformation is stored in only one result column. The formula may contain one or more references to other columns from which the result column will be calculated. The formula must not reference the result column. The result column is always locked when a transformation is active and cannot be edited except by changing values in columns included in the transformation formula. When you change the name of a column that is included in a transformation function, the function will be automatically updated with the new column name.

Transformation Editors

The Transformation Editor windows may be used to facilitate the entry of transformation formulas. You can think of these windows as special viewers for the Transformation column of the Column Info Table. The transformation editors are particularly useful when you don't know which function to use or when creating conditional and recode transformations. Remember that you do not have to use the transformation editors to create transformations; you can type any transformation formula directly into the Column Info Table.

There are several ways to launch the transformation editors.

1. From the Data Window main toolbar, click on the *Trans* button.

2. From the Data Window transformation toolbar, click on the *Transformation* button.

3. From the Data Window menu, select *Data > Transformations* and then one of the transformation types. You will be prompted to select a column in which to store the transformation.

4. From the Data Window menu, select *Data > Column X Info > Transformation* to create/edit a transformation for the column containing the active cell.

5. Click on the button at the right side of a transformation cell when it is selected in the Column Info Table to create/edit a transformation for the column containing the active cell.

6. Select a cell in the Data Table, right-click, and select *Column X Info > Edit Transformation* from the popup menu.

This editor that is launched depends on the type of transformation in the active column. If there is no transformation formula for the selected column, the general transformation editor will appear.

Calculation

The calculation of transformations is executed depending on the *Auto Recalculate Transformations* option located on the Transformation Toolbar (To view the Transformation Toolbar, select *View > Transformation Toolbar* from the Data Window menu). If the option is checked then transformations will be applied automatically (this is the default setting); otherwise, transformations will only be applied when the user clicks the *Recalculate All Transformations* button. When *Auto Recalculate Transformations* is on, all transformations will be recalculated every time one of the columns included in the transformations is changed. When *Auto Recalculate Transformations* is turned off, the user will be prompted to recalculate transformations whenever a column that is involved in a transformation is edited (the column background in the Data Table will become light red and a *Recalc All Transformations* button will appear on the bottom bar).

Order of Calculation

When transformations are calculated, the program will determine the correct order in which to execute transformations so that any result columns that are involved in other formulas will be calculated before downstream formulas. For example, if C3 contains the transformation "C1+C2" and C4 contains the transformation "C3*100", then the transformation in C3 should be executed before C4 because C4 depends on C3. The transformations do not have to be in any specific order going left to right for the calculations to happen correctly; **NCSS** figures out the correct order for you.

Activating and Deactivating Transformations

When you enter a transformation into a column, that column will become locked to data editing, and you will not be able to change data for that column in the Data Table without first deactivating the transformation. You will know that a column contains a transformation and is locked by the function-lock icon that appears in the column header of the Data Table. The background of the column in the Data Table will also become light blue. You can deactivate a transformation without deleting it by loading the transformation editor window and checking *Deactivate Transformation* or by simply typing a single quote character in front of the transformation in the Column Info Table (e.g. 'C1+C2). Deactivating the transformation will unlock the result column for editing. Inactive transformations appear grayed out. The transformation can be reactivated at any time by loading transformation editor window and unchecking *Deactivate Transformation* or by removing the single quote mark from the beginning of the transformation in the Column Info Table. When a transformation is reactivated, any data in the column containing the transformation will be overwritten, so be careful.

Note

The next item in the Column Info Table is the *Note* column. Column notes are stored with the dataset. You can use this space to record information about each column, save transformations or filters, or write comments. The notes are not used in any reports or plots.

Filter

The next item in the Column Info Table is the *Filter* column. For more information about filters and specific syntax, see the *Filters* chapter.

Filters are conditional statements about data in columns that allow you to restrict analysis and graphics to only 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*. For more information about filters and specific syntax, see the *Filters* chapter.

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>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

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 both 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 toos 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 main toolbar, click on the *Filter* button.

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 on the Data Window main toolbar controls whether or not 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* option on the Data Window main toolbar determines whether or not 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.

Advanced Filter (Not Always Displayed)

The final item in the Column Info Table is the *Advanced Filter* statement. This is only displayed if an advanced filter statement has been entered in the Filter Editor. Once an advanced filter statement has been created with the Filter Editor, you can edit it directly on the Column Info Editor. An advanced filter statement is not restricted to involve only one column but may involve multiple columns in a single statement. For more information, see the *Filters* chapter.

The Summary Statistics Table

The *Summary Statistics Table* provides user-selected statistical summaries of the data in each column. The statistics are calculated column by column, i.e. the calculations for one column are not influenced by data in another column (except in the case where a filter is running and rows are excluded based on values in other columns). The statistics are updated each time a change is made to the data.

Click *Statistics & Formatting* on the right side of the Summary Statistics Table title bar to select from over 40 available summary statistics, to change the order in which statistics are displayed, and to adjust the display formatting.

The entire summary table (including row and column headings) can be copied to the clipboard by right-clicking on the Summary Statistics Table and selecting *Copy Table* from the menu that appears.

Summary Statistics

The following summary statistics are available in the table. Click *Statistics & Formatting* on the right side of the Summary Statistics Table title bar to select from over 40 available summary statistics.

Dataset Summaries

Last Row with Data

This is the last row *within the column* that has non-missing data.

Number of Empty Cells ("*Empty Cells*" in the Table)

The number of missing data values encountered *up to the last row with data in the entire database*. Missing values in a column after the last row with data within a column are included in this count.

Number of Cells with Data ("*Cells with Data*" in the Table)

The number of non-missing data values, *n*. These values may be either numeric or text.

Number of Unique Values ("*Unique Values*" in the Table)

The number if unique data values.

First Sorted Data Value ("*First Sorted Value*" in the Table)

The first value after all values in the column are sorted alpha-numerically.

Last Sorted Data Value ("*Last Sorted Value*" in the Table)

The last value after all values in the column are sorted alpha-numerically.

Data List Sorted Alpha-Numerically with Counts ("*Data List (Count)*" in the Table)

A sorted list of all data values in the column. The count for each value is given in parentheses. For example, the output *A (40); B (60)* would indicate that the unique values in the column are *A* and *B* and that there are *40 A's* and *60 B's*. This is commonly known as a frequency table.

Data List Sorted Alpha-Numerically with Percents ("*Data List (Percent)*" in the Table)

A sorted list of all data values in the column. The percentage of each value is given in parentheses. For example, the output *A (40%); B (60%)* would indicate that the unique values in the column are *A* and *B* and that *40 percent of values are A's* and *60 percent of values are B's*. This is commonly known as a frequency table.

Data List Sorted by Count ("*Data List Sorted by (Count)*" in the Table)

A list of all data values in the column, arranged by value count. The value(s) with the highest count are listed first, followed by the value with the second-highest count and so on. The count for each value is given in parentheses. For example, the output *B (60); A, C (40)* would indicate that the unique values in the column are *A*, *B*, and *C* and that there are and *60 B's*, *40 A's*, and *40 C's*. This is commonly known as a frequency table.

Data List Sorted by Percent ("*Data List Sorted by (Percent)*" in the Table)

A list of all data values in the column, arranged by value percent. The value(s) with the highest percent are listed first, followed by the value with the second-highest percent and so on. The percentage of each value is given in parentheses. For example, the output *B (50%); A, C (25%)* would indicate that the unique values in the column are *A*, *B*, and *C* and that *50 percent of values are B's*, *25 percent of values are A's*, and *25 percent of values are C's*. This is commonly known as a frequency table.

Number of Numeric Values ("*Numeric Values*" in the Table)

This is the number of values in the column that are numeric. (Date & Time values are considered to be numeric in this context if the column data type is DateTime.)

Numeric Data Summaries

The summary statistics in this section are only calculated for columns with at least one numeric data value. (For columns with the DateTime data type, only the Minimum and Maximum are calculated.)

Sum

The sum (or total) of the data values.

$$Sum = \sum_{i=1}^{n} x_i$$

Mean

The average of the data values.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Median

The median (or 50th percentile) is the "middle number" of the sorted data values.

$$Median = Z_{50}$$

Mode(s) ("*Mode(s) (Count)*" in the Table)

The most frequently occurring value(s) is/are listed. The associated value count is given in parentheses.

Lower and Upper 95% Confidence Limits for the Mean ("*Lower 95% CL Mean*" and *Upper 95% CL Mean*" in the Table)

This is the upper and lower values of a 95% confidence interval estimate for the mean based on a *t* distribution with $n - 1$ degrees of freedom. This interval estimate assumes that the population standard deviation is not known and that the data for this variable are normally distributed.

$$95\% \text{ CI} = \bar{x} \pm t_{a/2, n-1} s_{\bar{x}}$$

Standard Deviation (SD) ("*SD*" in the Table)

The sample standard deviation, *s*, is a popular measure of dispersion. It measures the average distance between a single observation and the mean. It is equal to the square root of the sample variance.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Standard Error (SE) ("*SE*" in the Table)

The standard error of the mean, a measure of the variation of the sample mean about the population mean, is computed by dividing the sample standard deviation by the square root of the sample size.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Variance

The sample variance, s^2, is a popular measure of dispersion. It is an average of the squared deviations from the mean.

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Shapiro-Wilk Normality Test P-Value ("*S-W Normality Test P-Value*" in the Table)

This test for normality has been found to be the most powerful test in most situations. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of *n* observations. The numerator is proportional to the square of the best linear estimator of the standard deviation. The denominator is the sum of squares of the observations about the sample mean. The test statistic *W* may be written as the square of the Pearson correlation coefficient between the ordered observations and a set of weights which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics, *W* is roughly a measure of the straightness of the normal quantile-quantile plot. Hence, the closer *W* is to one, the more normal the sample is.

The test was developed by Shapiro and Wilk (1965) for samples up to 20. **NCSS** uses the approximations suggested by Royston (1992) and Royston (1995) which allow unlimited sample sizes. Note that Royston only checked the results for sample sizes up to 5000, but indicated that he saw no reason larger sample sizes should not work.

The probability values for W are valid for samples greater than 3. The test may not be as powerful as other tests when ties occur in your data.

Minimum

The smallest numeric data value. (Also calculated for the DateTime column data type.)

Maximum

The largest numeric data value. (Also calculated for the DateTime column data type.)

Range

The difference between the largest and smallest data values.

$$Range = Maximum - Minimum$$

Interquartile Range (IQR) ("*IQR*" in the Table)

The difference between the 75th and 25th percentiles (see below, the 3rd and 1st quartiles). This represents the range of the middle 50% of the data. It serves as a robust measure of the variation in the data.

$$IQR = Z_{75} - Z_{25}$$

Percentiles (1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, 99th) ("*[P]th Percentile*" in the Table)

The $100p^{th}$ percentile is the value below which $100p\%$ of data values may be found (and above which $100p\%$ of data values may be found). The $100p^{th}$ percentile is computed as

$$Z_{100p} = (1-g)X_{[k1]} + gX_{[k2]}$$

where $k1$ equals the integer part of $p(n+1)$, $k2=k1+1$, g is the fractional part of $p(n+1)$, and $X_{[k]}$ is the kth observation when the data are sorted from lowest to highest.

Mean Absolute Deviation from the Mean (MAD) ("*MAD*" in the Table)

A measure of dispersion that is not affected by outliers as much as the standard deviation and variance. It measures the average absolute distance between a single observation and the mean.

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Mean Absolute Deviation from the Median (MADM) ("*MADM*" in the Table)

A measure of dispersion that is even more robust to outliers than the mean absolute deviation (MAD) since the median is used as the center point of the distribution. It measures the average absolute distance between a single observation and the median.

$$MADM = \frac{\sum_{i=1}^{n} |x_i - Median|}{n}$$

Coefficient of Variation (COV) ("*COV*" in the Table)

A relative measure of dispersion used to compare the amount of variation in two samples. It is calculated by dividing the standard deviation by the mean. Sometimes it is referred to as COV or CV.

$$COV = \frac{s}{\bar{x}}$$

Coefficient of Dispersion (COD) ("*COD*" in the Table)

A robust, relative measure of dispersion. It is calculated by dividing the robust mean absolute deviation from the median (MADM) by the median. It is frequently used in real estate or tax assessment applications.

$$COD = \frac{MADM}{Median} = \frac{\left(\dfrac{\sum_{i=1}^{n} |x_i - Median|}{n}\right)}{Median}$$

Skewness

Measures the direction and degree of asymmetry in the data distribution.

$$Skewness = \frac{m_3}{m_2^{3/2}}$$

where

$$m_r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^r}{n}$$

Kurtosis

Measures the heaviness of the tails in the data distribution.

$$Kurtosis = \frac{m_4}{m_2^2}$$

where

$$m_r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^r}{n}$$

Possible Outliers (α = 0.05) ("*Possible Outliers (Row)*" in the Table)

The results of Grubbs' Outlier Test are presented. Only values for which a two-sided test is significant at alpha = 0.05 are given. Each "possible" outlier is listed followed by the row on which it is found in parentheses. If no "possible" outliers are found, the result is blank.

Grubbs' Outlier Test requires the user to specify how the maximum number of possible outliers is calculated. These options are accessible on the Summary Statistics Table Options window that can be accessed by clicking on *Statistics & Formatting*.

For more information by Grubbs' Outlier Test, see chapter 203 in the documentation.

Section Separators (----- [Blank 1] -----, ----- [Blank 2] -----, ----- [Blank 3] -----)

These options provide the ability to insert blank rows to separate various statistics. The blanks can be moved using the up and down arrows on the Summary Statistics Table Options window that can be accessed by clicking on *Statistics & Formatting*.

Menus

We will discuss the various options that are on the Data Window menus.

File Menu

The File Menu controls the opening and closing of datasets, importing, exporting, and printing. Note that some of the basic File Menu operations are also provided on the main toolbar.

We will now discuss each of these options.

- **New**

  This option closes the current dataset (if any) and creates a new dataset. If you have made changes to the current dataset, you will be prompted to save the dataset before it is cleared.

- **Open and Recent**

  These options let you open existing **NCSS** datasets. The opened dataset is copied into memory. Once you open a dataset, you actually have two copies of it—one in memory and one on your disk. No automatic relationship is maintained between the loaded dataset and the disk dataset. Changes made to the copy in memory will <u>not</u> automatically change your disk data files unless you save them!

- **Open Example Data**

  This option takes you directly to the location on your computer where you chose to install **NCSS** so you can quickly open an example data file.

- **Save**

  This option saves the current dataset.

- **Save As**

  This option saves the current dataset to a data file with a different name. For example, if you are working on a dataset called "XYZ" and will be making changes to it, you might want to save a copy of it as "XYZ1" so that any mistakes you might have made will not destroy your original data.

- **Save Subset As**

  This option allows you to save a subset of the current dataset. You can save a subset of rows and/or columns by entering a list of rows and columns, or you can save a subset of rows using the active filter.

- **Import**

  This option lets you import data from various other spreadsheets, databases, text files, and statistical systems into **NCSS**. For more information, see the *Importing Data* chapter.

- **Export**

  This option lets you export the current database to another format for use in other spreadsheets, databases, etc. For more information, see the *Exporting Data* chapter.

- **Export Subset As**

  This option exports the current dataset using the current filter. Only the values that pass the data filter are saved to the exported file. This is particularly useful for exporting data subsets.

- **Print, Print Preview, Print Setup**

  These options allow you to setup and print the Data Table.

- **Dataset Notes**

  This option brings up a text editor window that will allow you to write notes pertinent to the dataset. Dataset notes are saved with the dataset.

- **System Options**

  This brings up the System Options window to set various system options.

- **Exit NCSS**

  This option quits the **NCSS** system. If your dataset has changed, you will be prompted to save it before exiting.

- **Previously Opened NCSS Files and Previously Imported Files**

  A list of previously opened datasets and imported files is presented. You may select any of them to revert to that dataset directly or automatically start an import. Mouse over the file name to view the complete path and file name.

Edit Menu

The Edit Menu controls the editing of data in the Column Info Table and the Data Table. Note that some of the basic Edit Menu operations are provided on the Toolbar. Many of the options should be familiar and will not be described in detail.

- **Undo/Redo**

  Undo/Redo allows you to undo (and redo) operations going back to the last time the dataset was opened or saved. Each time the dataset is saved, the undo stack is cleared. The undo/redo operations are recorded from both the Column Info Table and the Data Table.

- **Copy with Filter**

  This option performs the same action as Copy, except that only values that pass the filter are copied. This option is only available when there is a filter active.

- **Copy Table**

  This option copies the entire contents of the currently-focused table to the clipboard, including row and column headers.

- **Copy Table with Filter**

  This option performs the same action as Copy Table, except that only rows that pass the filter are copied. This option is only available when there is a filter active.

- **Paste Rotated**

  This option works like Paste, except that the data are rotated ninety degrees so that the rows become the columns and the columns become the rows.

  The following example shows the result of using this option:

  Copied Data
1	2
3	4
5	6
7	8

Paste Rotated Result

1	3	5	7
2	4	6	8

- **Clear**

   This option clears the contents of the selected cells. This does not delete the rows or columns, just the values in the cells.

- **Insert Row(s) or Column(s)**

   This option allows to you insert rows or columns at the current location.

- **Delete Row(s) or Column(s)**

   This option allows to you delete rows or columns at the current location.

- **Add Rows/Columns**

   This option allows to you add rows or columns to the end of the dataset. This can be done at any time.

- **Resize Rows/Columns**

   These options allow you to resize the columns and rows in the entire database all at once to defaults, to fit the data and titles in each row and column, or to fit just the data in each row and column. You can resize columns or rows individually by placing the mouse cursor over the border in column or row header area and then left-clicking and dragging the cursor.

View Menu

The View Menu controls what objects are displayed on the Data Window. Using this menu you can choose to display value labels on the Data table. You can also enlarge the tables using the zoom options.

Data Menu

The Data Menu lets you select tools and options that modify the dataset itself.

- **Rename Column X**

   This brings up the Column Name/Label Editor window.

- **Column X Info**

   This submenu allows you to bring up the editor window for each Column Info Table item. The editor comes up ready to edit the currently-active column.

- **Column X Font**

   This option allows you to change the display font of the currently-active column.

- **Filter**

   Contains options to edit the filter, activate/deactivate the filter, and hide/show filtered rows.

- **Transformations**

   Contains options to edit transformations, turn Auto Recalculate Transformations On and Off, and recalculate all transformations. It also has options for Box-Cox transformations.

- **Value Labels**

  Contains an option that allows you to load the Value Label List Manager to view and edit available value label lists associated with the dataset.

- **Sort**

  This option brings up a window that lets you select columns to sort by. The entire Data Table will be sorted.

- **Fill**

  This option brings up a window that allows you to fill empty columns with data series, sequences, or random uniform and normal numbers.

- **Stack Data**

  This option brings up a window that allows you to stack columns and store the result in a new column.

- **Unstack Data**

  This option brings up a window that allows you to unstack a column and store the results in multiple columns.

- **Create Contrast Variables**

  This option brings up a window that allows you to create binary or contrast variables from a single column with categorical data.

- **Various Data-Related Procedures by Name**

  Contains procedures associated with modifying, screening, matching, summarizing, and merging datasets.

Analysis and Graphics Menus

These menus load the corresponding procedure windows.

Tools Menu

From this menu you can load various statistical calculators, data procedures, and the macro command center window. You can also play the active macro from this menu.

Window Menu

This menu lets you transfer to one of the other **NCSS** menus such as the Output window or one of the currently open procedure windows.

Help Menu

From this menu you can launch the **NCSS** Help System and view documentation, tutorials, and references. From this menu you can also view licensing information.

Toolbars

The toolbars are provided for single-click access to the most commonly used menu options. You will find that each of the options on the toolbar can also be found in the menus. The *Main Toolbar* is located at the top of the screen just under the menus. On the left, the main toolbar contains buttons and options related to the Data Window. On the right, the main toolbar contains navigation buttons to help you quickly move among windows in the system and load procedures. The *Transformation Toolbar* (when showing) is located just under the main toolbar. By default, the transformation toolbar is hidden. To show the transformation toolbar select *View > Transformation Toolbar* from the menu. The *Bottom Bar* is located at the bottom of the screen. The bottom bar contains a current row/column indicator at the left that lets the user know where the currently-active cell is located. The bottom bar is frequently updated with operation status messages when data procedures are run. The bottom bar will also contain information about the current filter and indicate the need for transformations to be calculated (when applicable). Finally, the bottom bar contains a tool at the right that lets you adjust the size of text that is displayed.

Data Entry Bar

The *Data Entry Bar* provides an alternate place to edit data. As you move around the Data Table, the contents of the active cell are copied to this Data Entry Bar. Occasionally, your data will be longer than can easily be displayed in the cell. Although you could reset the column width, you usually find it easier to edit the data in the Data Entry Bar. Any changes you make will not be entered into the Data Table until you hit the ENTER key, TAB key, or position the mouse on another cell.

Press the expand/collapse button located to the right of the Data Entry Bar to change the size of the bar. Or, put your mouse pointed over the horizontal bar just below the Data Entry Bar, left-click, and drag to resize the bar.

Navigating the Data, Column Info, and Summary Statistics Tables

This section describes how to move around in and edit the tables using the keyboard and the mouse. The Summary Statistics Table is not editable, but you can copy values from the table to the clipboard.

Active Cell

The cursor is always located on a single cell, even when a range of cells is selected. The cell on which the cursor is located is called the *Active Cell*. Any typing that is done will only affect the active cell. The contents of the active cell are displayed in the *Data Entry Bar*. The address of the active cell is displayed in the bottom bar.

Keyboard Commands

The following commands are used for data entry and for moving about the tables.

Table of Keyboard Commands

Key	Description
Enter	Accepts the current entry and moves the active cell down one cell. When a range of cells is selected, accepts the current entry and moves down to the next selected cell. When the bottom of the selection is reached, the active cell moves to the top of the next selected column to the right.
Shift + Enter	Acts like the Enter key, except that cell-to-cell movement is upward and to the left instead of downward and to the right.
Tab	Accepts the current data entry and moves the cursor one cell to the right. When a range of cells is selected, the tab moves the cursor to the right to the next cell in the selection.
Shift + Tab	Acts just like the TAB key, except that cell-to-cell movement is to the left instead of to the right.
F2	Enters edit mode.
Delete	Clears the current entry or selection.
Esc	Cancels the current data entry.
Up/Down/Left/Right Arrow	Moves the active cell Up/Down/Left/Right one row or column.
Ctrl + Up/Down/Left/Right	Moves to the next range of cells containing data. If there is no additional data in any of the cells in that direction, the active cell is moved to the edge of the Data Table.
Page Up	Moves up one screen.
Page Down	Moves down one screen.
Ctrl + Page Up	Moves left one screen.
Ctrl + Page Down	Moves right one screen.
Home	Moves to the first column in the current row.
End	Moves to the last column in the current row that contains data.
Ctrl + Home	Moves to the upper-left corner of the Data Table (cell 1,1).
Ctrl + End	Moves to the last row and column that contains data.
Shift + Movement Key	Extends the current selection in the direction indicated.

Mouse Actions

The mouse is used mainly for positioning the active cell and making selections. You can also use the mouse to move a block of cells around the Data Table.

Table of Mouse Commands

Key	Description
Left Click	Accepts the current entry and moves the active cell to the position of the mouse.
Right Click	Loads the edit popup menu.
Left Click in a Row or Column Heading	Accepts the current entry and moves the active cell to the position of the mouse.
Left Double-Click	In-cell editing is invoked.
Right Double-Click	Does nothing.
Left Click and Drag	Selects a range of cells. If other ranges were selected, they are unselected.
CTRL + Left Click and Drag	Selects a range of cells. If other ranges were selected, they remain selected. Note that edit commands, such as cut and copy, will only work on a single range.
SHIFT + Left Click and Drag	Extends the current selection in the direction indicated.
Dragging a Selection's Copy Handle	Copies the selection to a new location. The copy handle is the small plus sign at the lower-right corner of a selection.

Selecting Cells

Many operations require one or more cells to be selected. There are three types of selections: a single cell, a single rectangular range of cells, and multiple ranges of non-adjacent cells. Cells may be selected either with the mouse or with the keyboard.

Selecting Cells with the Mouse

To select a range of cells with the mouse, click and hold the left mouse button down on the upper-left cell of the range you want to select. Drag the mouse cursor to the lower-right cell, while continuing to hold the left mouse button down. When the desired cells are selected, release the mouse button.

To select multiple ranges with the mouse, press the CTRL key while making each additional selection. Note that multiple selections are only useful for controlling cell cursor movement during data entry. They are not used by any of the edit functions.

To select an entire row or column, click on the row or column heading.

Once a range is selected, you can move the active cell within the selection using the Enter, SHIFT+ENTER, TAB, and SHIFT+TAB keys without destroying the selection.

Selecting Cells with the Keyboard

To select a range of cells with the keyboard, position the active cell at the upper-left corner of your desired selection. While holding down the SHIFT key, use the cursor movement keys (such as the arrow keys) to move to the lower-right corner of the selection.

Editing Tables Interactively

Data can be entered into the tables in many different ways. In this section, we will explain how you can quickly move and copy ranges of cells by clicking and dragging the copy handle of a selection.

Copying Data Interactively

You can copy a range of cells quickly by using only a few clicks of your mouse.

1. Select a range of one or more cells.

2. Position the mouse pointer on the lower-right corner of the selection (the *copy handle*). When positioned on the copy handle, the pointer changes to cross-hairs.

3. Left-click and drag the copy handle through the range of cells that are to receive the copied data.

4. Release the left mouse button.

Moving Data Interactively

You can move a range of cells quickly by using only a few clicks of your mouse.

1. Select a range of one or more cells.

2. Position the mouse pointer on the border of the selection. When positioned on the border, the pointer changes to an arrow.

3. Left-click and drag the selection to the new location.

4. Release the left mouse button