Welcome to Statcato

Statcato is a Java application for elementary statistics computations. Its features include manipulating and generating data; computing probability distributions, descriptive statistics, confidence intervals; performing hypothesis tests, correlation and regression, analysis of variance; and generating simple graphs.

Getting Started

Statcato has a log window that displays outputs of computations and a datasheet window where user data can be provided.

Various utilities are provided in the menu bar, separated into the following categories:

- **File**: Open new or saved projects/datasheets, clear or close windows, load datasets, save and print files, options
- **Edit**: Cut, copy, paste, select all, clear, delete, insert rows, columns, or cells into datasheet, edit last dialog, show dialog history
- **Data**: Sort, rank, standardize data, display datasheet data in log window, generate patterned or random data
- **Calculate**: Compute probability distributions, p-Values, and frequency table; calculator; hypothesis test conclusion tool
- **Statistics**: Descriptive statistics, sample size determination, confidence intervals, hypothesis tests, correlation and regression, multinomial experiments, and analysis of variance
- **Graph**: Bar chart, box plot, dot plot, histogram, normal quantile plot, pie chart, scatterplot, stem-and-leaf plot
- **Help**: Help menu, program information, check for updates

A toolbar providing shortcuts to common operations is available below the menu bar.

Introduction > Toolbar

The toolbar at the top of the application window provides shortcuts to commonly used functions.

1. **New Datasheet**: Open a new datasheet in the Datasheet window.
2. **Open**: Open a saved log if the log window is selected. Open an existing Datasheet if the datasheet window is selected.
3. **Save**: Save the log window if the log window is selected. Save the datasheet if the datasheet window is selected.
4. **Import**: Load a built-in datset or one from a web address.
5. **Print**: Print the log window if it is selected. Print the active datasheet if the datasheet window is selected.
6. **Cut**: Cut.
7. **Copy**: Copy.
8. **Paste**: Paste.
9. **Undo**: Undo.
10. **Redo**: Redo.
11. **Edit Last Dialog**: Bring up the most recently used dialog.
12. **Dialog History**: Show a history of the most recently used dialog.
13. **Help**: Display the documentation.

Introduction > Keyboard Shortcuts

File

New Datasheet
    Ctrl-N
Open Datasheet
    Ctrl-O
Save Current Datasheet
    Ctrl-S
Print Log Window
    Ctrl-P
Exit
    Alt-F4

Edit

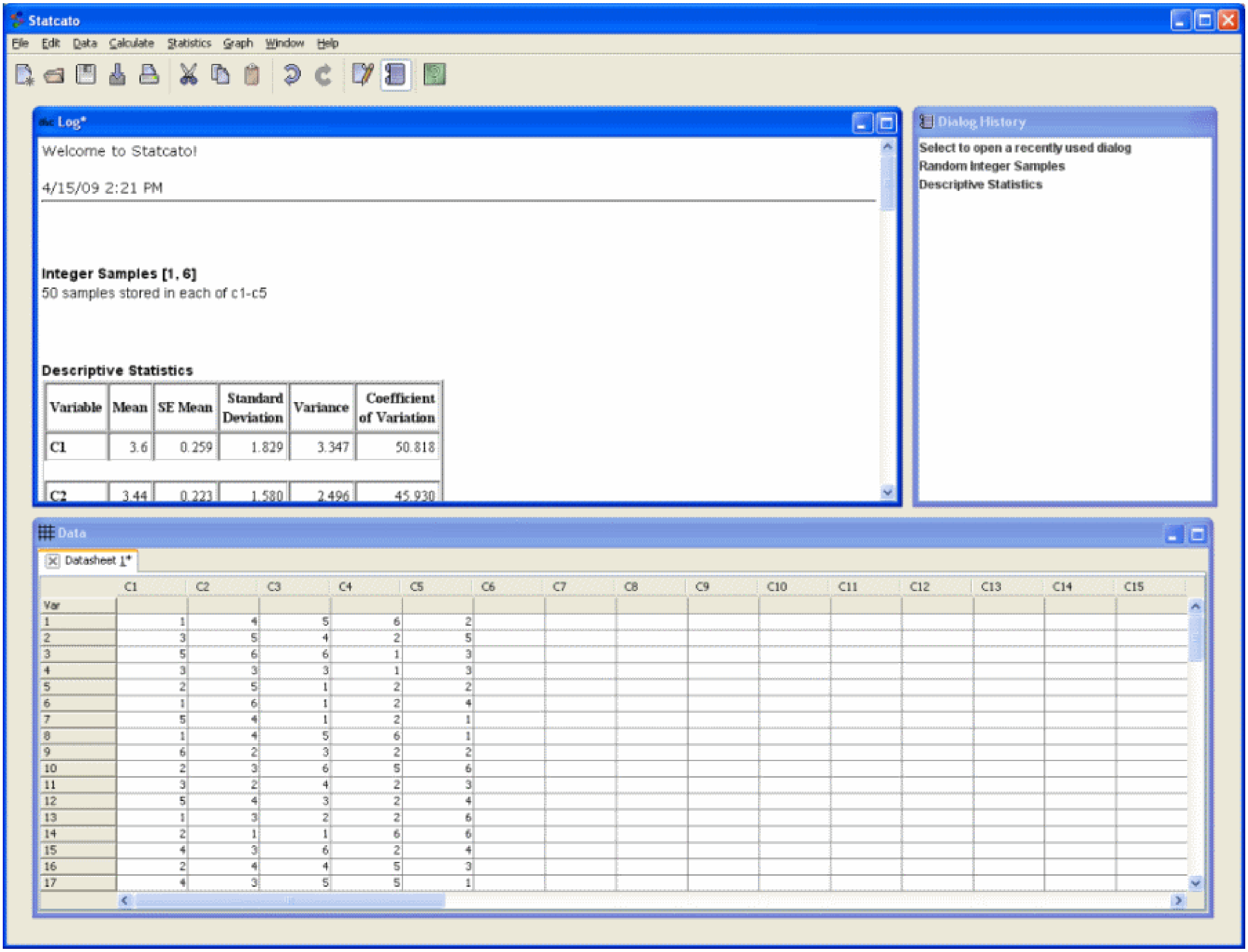
Undo
    Ctrl-Z
Redo
    Ctrl-Y
Cut
    Ctrl-X
Copy
    Ctrl-C
Paste
    Ctrl-V
Select All
    Ctrl-A
Clear
    Backspace
Delete
    Delete
Edit Last Dialog
    Ctrl-D

Help

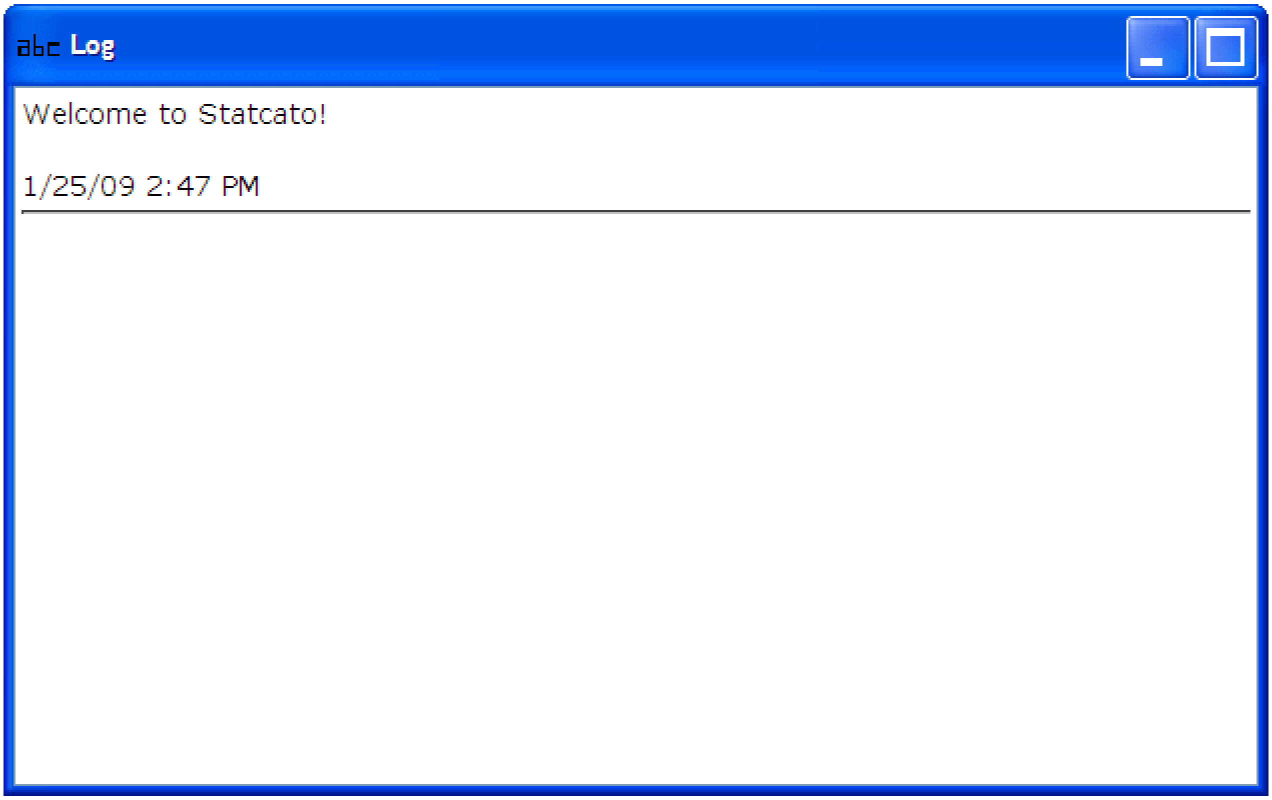
Help Menu
    F1
About
    F2

Windows

Statcato has a Log window that displays outputs of computations, a Data window where user data can be provided, and a Dialog History window that provides access to recently used dialogs.



Log Window



The log window is a window where the output of various commands of the application is displayed. It is an internal window located at the top half of the Statcato main window. The contents of the window is displayed using HTML.

Editing the log window

You can type in the log window to enter plain text. You can also perform common edit operations on the log window contents, such as copy (ctrl-c), paste (ctrl-v), and select all (ctrl-a).

Clearing the log window

To clear the log window, select **Clear Log** from the **File** menu.

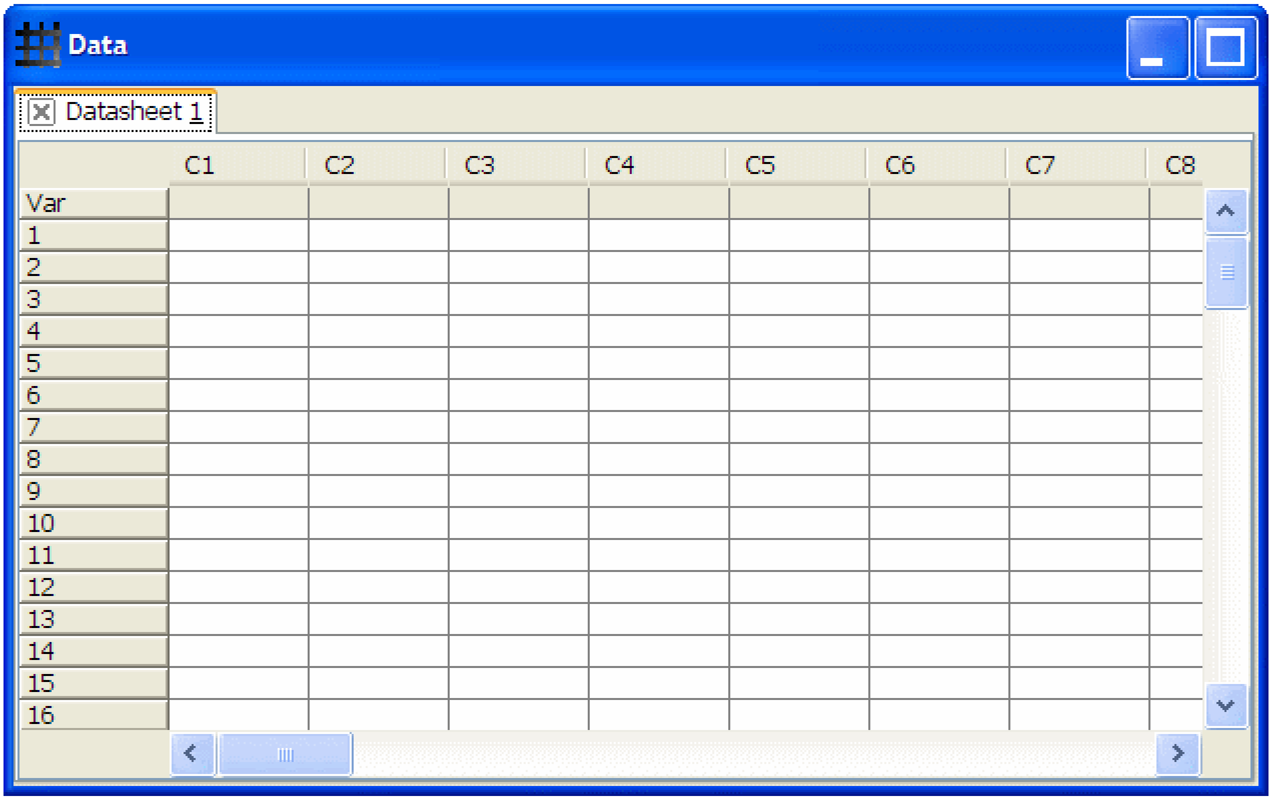
Printing the log window

Select **File > Print Log ...** (ctrl-p). A Print Log Window dialog will appear and provide the following print options:

- Header: Check the header checkbox to print a header at the top of the document. Enter the header in the provided text box.
- Footer: Check the footer checkbox to print a footer at the bottom of the document. Enter the footer in the provided text box. The page number is used in the footer by default (Page {0}).

Press Print to print the log window.

Datasheet Window



The Datasheet Window contains Datasheets where you can enter data. A Datasheet takes the form of a spreadsheet and is placed within a tab. By default, a Datasheet has 300 rows (1, 2, ..., 300) and 50 columns (C1, C2, ..., C50).

The **variable** row of the table is for the variable names of the columns. Variables names cannot contain characters ' and ", and they cannot be the same for different columns.

A cell in a datasheet can contain both text and numbers. Numbers in data cells are internally stored as double-precision 64-bit IEEE 754 floating point values, even though they are displayed to at most 6 decimal places. The tooltip of a cell shows the precise value stored in a cell.

Adding a New Datasheet

To add a new Datasheet to the Datasheet window, select **File > New Datasheet**.

Saving a Datasheet

A Datasheet can be saved in the following formats:

- Tab-delimited values (.txt)
- Comma-separated values (.csv)
- Excel (.xls)

To save a Datasheet, make sure it is active (by selecting its corresponding tab) and select **File > Save Current Datasheet...**. In the save file dialog, enter the filename and choose the desired file format. To save the current Datasheet as a different file, use **File > Save Current Datasheet As...**.

Opening a Datasheet / Import Data

To open a saved Datasheet or import data from a file, select **File > Open Datasheet**. The following file formats are supported:

- Tab-delimited values (.txt)
- Comma-separated values (.csv)
- Excel (.xls)

Closing a Datasheet

Before closing a Datasheet, make sure it is the current Datasheet by selecting its corresponding tab. You can close a Datasheet by one of the following two ways:

- Select **File > Close Current Datasheet**.
- Click the cross button next to the Datasheet name on the tab.

You will be given the option to save the Datasheet before closing it if there is any unsaved data.

Printing / Exporting Data

To print a Datasheet or export data to a file, select **File > Print/Export Datasheet**. In the print Datasheet dialog, select the desired print area:

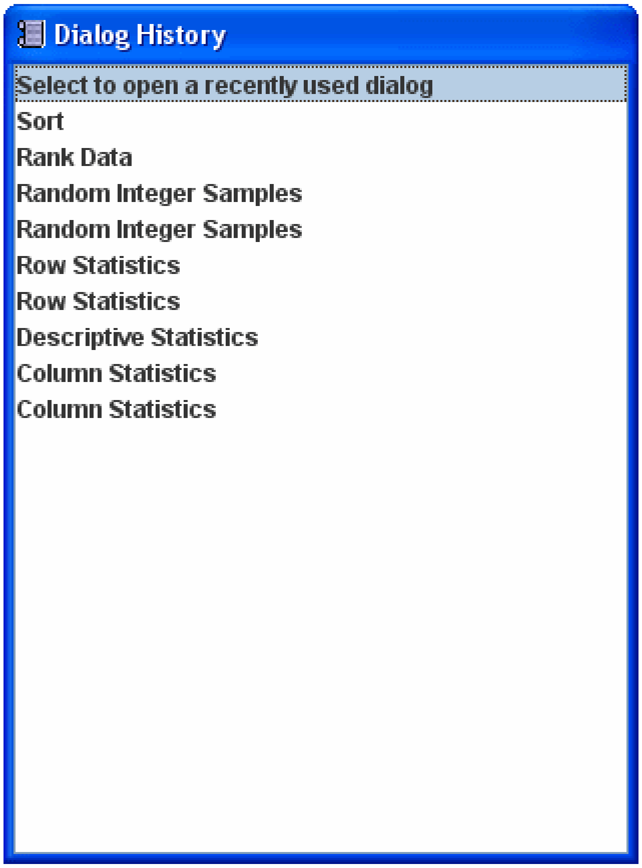
- All: The entire Datasheet
- Selection: The selected (highlighted) cells
- Specified rows and columns: Enter the starting and ending row/column numbers

Enter in the given text boxes the desired header and footer to be printed. Press **Print** to continue the printing or exporting process.

A Print Preview window will appear. Use the File Menu to export data to the following formats:

- PDF (Portable Document Format)
- Excel
- RTF (Rich Text Format)
- HTML (HyperText Markup Language)
- CSV (Comma-Separated Values)
- Text file

Dialog History Window



The dialog history window shows the 20 most recently used dialogs. Selecting an item in this window brings up the corresponding dialog.

Windows > Menu

The windows menu provides access to the internal and external windows within the application. It contains menu items for the log, datasheet, and history windows, as well as for each chart frames created by Statcato. Selecting a window menu item brings the corresponding window to the front.

File

Common operations for manipulating files in Statcato can be accessed from the File menu.

File

A Statcato project file contains a log and a set of datasheets. You can save and open project files using the menu items in the file menu. You can also open and save individual datasheets from the file menu.

New Datasheet

Inserts a new Datasheet in the Datasheet window.

Open Project

Opens an existing Statcato project file (.stc).

Save Project

Saves the current application session (log, datsheets, and graphs) to a Statcato project file.

Save Project As

Saves the current application session (log, datsheets, and graphs) as a Statcato project file.

Close Project

Closes the current project file.

Open Datasheet

Opens a datasheet in the Datasheet window. Accepted Datasheet formats include tab-delimited text, comma-separated values, and Excel files.

Save Current Datasheet

Saves the current datasheet in tab-delimited text, comma-separated values, or Excel file formats.

Save Current Datasheet As

Saves the current datasheet as a specific file.

Close Current Datasheet

Closes the current Datasheet.

Load Dataset

Load a built-in sample dataset or a dataset from the web by specifying its web address. Supported dataset formats are tab-delimited text, comma-separated values, and Excel files.

Clear Log

Clears the contents in the log window.

Print Log

Print the log window.

Print/Export Datasheet

Print the current Datasheet or export it to a file.

Options

Opens the options dialog.

File > Options

This dialog allows the user to set options of this application.

Auto Save

- Enter the interval in which a back up file of the current project file is saved. The interval must be a positive integer and is measured in minutes.
- Select the **Delete backup file upon exit** checkbox if you want the backup file to be deleted when you exit the application.

Auto Save Feature

The program automatically saves a back up copy of the current project file at one minute after the start of the program execution. The back up file is saved in the same directory as the program files and is named in the form ~statcato.tmp.mmddyyyyhhmmss.stc, where mmddyyyyhhmmss is a timestamp of when the file is created.

By default, auto save happens every five minutes. You can change the frequency by going to File > Options.

By default, the back up file is retained after the application exits. If you would like the back up file to be deleted after the application exits, you can change the auto save options by going to File > Options.

Edit Data

Editing functions can be accessed from the edit menu.

Edit Menu

The edit menu provides functions for editing Datasheets.

- **Cut**: Cut selected contents to the clipboard.
- **Copy**: Copy selected contents to the clipboard.
- **Paste**: Paste contents from the clipboard to the insertion point.
- **Select All**: Select all cells in the current Datasheet
- **Clear Cells**: Clear the contents in the selected cells
- **Delete Cells**: Delete the selected cells. Cells below the deleted cells are shifted up to replace the deleted ones. The size of the Datasheet is kept unchanged.
- **Insert Rows**: Choose **Above** to insert a blank row above the current cell. Choose **Below** to insert a blank row below the current cell.
- **Insert Columns**: Choose **Left** to insert a blank column to the left of the current cell. Choose **Right** to insert a blank column to the right of the current cell.
- **Insert Cell**: Choose **Above** to insert a blank cell above the current cell. Choose **Below** to insert a blank cell below the curent cell. Cells are shifted up/down to accommodate the inserted cell.
- **Add Multiple Rows/Columns**: Add a user-specified number of rows and columns to the end of the spreadsheet.

Data

The data menu provides utilities for manipulating and generating data in the Datasheets.

Data > Sort

The sort utility sorts a number of data columns by data up to four columns. You can store the sorted data in its original columns or in a specific set of columns.

If a column contains only numbers, the data will be sorted by their numerical values. Otherwise, the data will be sorted as text in alphabetically order.

- To open the sort utility, select **Data > Sort**.
- In the **Input Columns:** text box, enter the names of the columns. Separate column names by spaces (e.g. c2 c4), and use a dash to indicate a range of columns (e.g. c10-c15).
- To store the sorted data in the original columns, select **Store in original columns**.
- To store the sorted data in a specific set of columns, enter the column names in the **Store in these columns:** text box. The number of input columns must be the same as the number of columns for storing results. The results of sorting the input column(s) will be stored in the store columns in order.
- To sort the input columns by data in another column, select the appropriate column label in the **Sort By Column:** drop-down menu. Choose **Ascending** to sort in ascending order, or **Descending** to sort in descending order. To sort by more columns, use the **Then By Column:** drop-down menus. You may sort by at most four columns.
- Click **OK** to sort.

Example

C1	C2	C3
a	2	5
b	2	7
c	2	3
d	2	2
e	1	9
f	1	8
g	1	7
h	1	6

The result of sorting C1 by C2 and C3 is

h
g
f
e
d
c
a
b

Data > Rank

The rank utility assigns ranks to numerical values in a column. The ranks are computed as follows:

- The values are first sorted in ascending order.
- The smallest value is assigned a rank of 1. The next smallest value in the sorted list is assigned a rank of 2, and so on.
- The rank of repeated values is the average of their corresponding positions in the sorted list of values.

To use the utility, select **Data > Rank**. In the rank dialog, select the column containing data to be ranked using the **Rank Data in:** dropdown menu. Enter the column or variable name for storing ranks in the **Store Ranks in:** text box.

Example

The following shows the ranks of the values in C1:

C1	ranks
2	1.0
5	2.0
7	3.5
7	3.5
8	5.0
9	6.0

Data > Transpose

The transpose utility converts a column of data to a row of data and vice versa.

To use the utility, select **Data > Transpose**. In the transpose dialog, select whether you want to transpose a column into a row, or transpose a row into a column. Enter the column or variable name as well as the row number in the corresponding text boxes.

Data > Display

This utility displays the contents of column variables, along with the row numbers, in a tabular format in the log window.

To use the utility, select **Data > Display Data in Log Window...**.

In the list of columns, select the column(s) to be displayed in the log window (control-click to select multiple columns). Click **OK** to display data.

Data > Standardize

The **Standardize** utility performs data transformation on columns of data, including scaling and moving the center of data in a column, as well as adjusting the range of the data.

To use the utility, select **Data > Standardize...**.

- In the **Input Column(s)** text box, enter the names of the columns. Separate column names by spaces (e.g. c2 c4), and use a dash to indicate a range of columns (e.g. c10-c15).
- In the **Store results in column(s)** text box, enter valid column names where the results will be stored. The number of input columns must be the same as the number of columns for storing results. The results of standardizing the input column(s) will be stored in the store columns in order.
- Select one of the following options:
  - **Subtract mean and divide by standard deviation**: For each number in the column, subtract the mean of the column data set and divide by the standard deviation.
  - **Subtract ... Then divide by ...**: For each number in the column, subtract the first specified number and then divide by the second specified number.
  - **Make range From: ... To: ...**: For each number in the column, shift and scale so that the range of the data goes from **From** to **To**.
- Click OK to finish.

Manipulate Data > Generate Patterned Data

Statcato provides utilities for generating two types of patterned data: (1) Simple number sequence, in which numbers differed by a common step size are repeated individually or as a sequence; and (2) arbitrary data sequence, in which arbitrary data values are repeated individually or as a sequence.

Manipulate Data > Generate Patterned Data > Simple Number Sequence

This utility generates a simple number sequence in a column. The sequence consists of repetitions of terms that are differed by a constant step size. To use this utility, you will specify the column for storing the sequence, the first and last value of the sequence, the step size, the number of times a value is repeated, and the number of times the whole sequence of values is repeated.

- The utility can be accessed from **Data > Generate Patterned Data > Simple Number Sequence**
- **Store number pattern in:** The column names in which the numbers will be stored, separated by spaces (e.g. c1 c3 c7). For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30)
- **From first number:** The first number in the sequence
- **To last number:** The last number in the sequence
- **In steps of:** The difference between two terms (step size)
- **Number of times to list each number**: The number of times each term appears in the sequence (continuously)
- **Number of times to list each sequence**: The number of times the sequence, from the first to the last value, is repeated

Example

Given the inputs (From first number: 1, To last number: 10, In steps of: 2, Number of times to list each number: 2, Number of times to list each sequence: 2) produces the following sequence: 1 1 3 3 5 5 7 7 9 9 1 1 3 3 5 5 7 7 9 9.

Manipulate Data > Generate Patterned Data > Arbitrary Data Sequence

This utility generates a arbitrary data sequence in a column. The sequence consists of repetitions of values and repetitions of the whole sequence itself. Values in the sequence can be numbers or text values. To use this utility, you will specify the column for storing the sequence, a data sequence, the number of times a value is repeated, and the number of times the whole sequence of values is repeated.

- The utility can be accessed from **Data > Generate Patterned Data > Arbitrary Data Sequence**
- **Store number pattern in:** The column names in which random samples will be stored, separated by spaces (e.g. c1 c3 c7). For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30)
- **Arbitrary data sequence**: a sequence of values separated by white space; for a text string with internal space characters, enclose text values with double quotes("")
- **Number of times to list each number**: The number of times each term appears in the sequence (sequentially)
- **Number of times to list each sequence**: The number of times the sequence, from the first to the last value, is repeated

Example

Given the inputs (Data sequence: "this is" 1 2 a b, Number of times to list each number: 2, Number of times to list each sequence: 2) produces the following sequence:
this is
this is
1
1
2
2
a
a
b
b
this is
this is
1
1
2
2
a
a
b
b

Data > Generate Random Data

This menu contains utilities for generating random data:

- Binomial: Generate samples from a binomial distribution given the number of trials and the event probability.
- Integer: Generate samples from a range of integers, each of which has the same probability of being chosen.
- Normal: Generate samples from a normal distribution.
- Uniform: Generate samples from a uniformly distributed range of continuous values.
- Samples from Column: Generate samples from a given column of data values. The option of random sampling with or without replacement is provided.

Data > Generate Random Data > Binomial

This utility simulates the process of drawing random samples from a binomial distribution. To use this utility, you must specify the column name(s) in which samples will be stored, the sample size, and the number of trials and the event probability of the binomial distribution. Providing a seed for the random number generator is optional.

Dialog Inputs

- In the **Store Samples in:** text field, enter the column names in which random samples will be stored, separated by spaces. For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30).
- In the **Number of Samples to Generate:** text box, enter the number of random samples to generate for each column.
- Enter the number of trials in the **Number of Trials:** text field.
- Enter the event probability in the **Event Probability:** text field.
- (optional) Enter a seed for the random number generator in the **Random Generator Seed:** text box.

Data > Generate Random Data > Integer

This utility simulates the process of drawing random samples from a range of integer values, each of which having the same probability of being chosen. To use this utility, you must specify the column name(s) in which samples will be stored, the sample size, and the minimum and maximum values of the range of possible values. Providing a seed for the random number generator is optional.

Dialog Inputs

- In the **Store Samples in:** text field, enter the column names in which random samples will be stored, separated by spaces. For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30).
- In the **Number of Samples to Generate:** text box, enter the number of random samples to generate for each column.
- Enter the minimum integer to generate in the **Minimum:** text field.
- Enter the maximum integer to generate in the **Maximum:** text field.
- (optional) Enter a seed for the random number generator in the **Random Generator Seed:** text box.

Data > Generate Random Data > Normal

This utility simulates the process of drawing random samples from a normal distribution. To use this utility, you must specify the column name(s) in which samples will be stored, the sample size, and the mean and standard deviation of the normal distribution. Providing a seed for the random number generator is optional.

Dialog Inputs

- In the **Store Samples in:** text field, enter the column names in which random samples will be stored, separated by spaces. For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30).
- In the **Number of Samples to Generate:** text box, enter the number of random samples to generate for each column.
- Enter the mean of the normal distribution in the **Mean:** text field.
- Enter the standard deviation in the **Standard Deviation:** text field.
- (optional) Enter a seed for the random number generator in the **Random Generator Seed:** text box.

Data > Generate Random Data > Uniform

This utility simulates the process of drawing random samples from a range of uniformly distributed continuous values. To use this utility, you must specify the column name(s) in which samples will be stored, the sample size, and the minimum and maximum values of the range of possible values. Providing a seed for the random number generator is optional.

Dialog Inputs

- In the **Store Samples in:** text field, enter the column names in which random samples will be stored, separated by spaces. For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30).
- In the **Number of Samples to Generate:** text box, enter the number of random samples to generate for each column.
- Enter the minimum value (inclusive) to generate in the **Minimum:** text field.
- Enter the maximum value (exclusive) to generate in the **Maximum:** text field.
- (optional) Enter a seed for the random number generator in the **Random Generator Seed:** text box.

Data > Generate Random Data > Samples from Colum

This utility simulates the process of drawing random samples from a set of data values with or without replacement. To use this utility, you must specify a column containing data values to draw samples from, column name(s) in which samples will be stored, whether the samples are replaced after being chosen, and the sample size. Providing a seed for the random number generator is optional.

Dialog Inputs

- Select the column containing data values to draw samples from in the **Samples from Column:** drop-down menu.
- In the **Store Samples in:** text field, enter the column names in which random samples will be stored, separated by spaces. For a continuous range of columns, specify using the first column, a dash, and the last column (e.g. C1-C30).
- Select the **Sample With Replacement** check box to sample with replacement. Deselect to sample without replacement.
- In the **Number of Samples to Generate:** text box, enter the number of random samples to generate. The sample size must be no larger than the number of possible data values if sampling without replacement.
- (optional) Enter a seed for the random number generator in the **Random Generator Seed:** text box.

Calculate Data

The **Calculate** menu provides utilities for manipulating and calculating data.

Calculate > Calculator

This utility parses and evaluates mathematical expressions. It supports addition, subtraction, multiplication, division, power, and a collection of common mathematical functions (such as sin, cos, tan, exponential, etc). Operands can be numbers or numerical data within columns, and the results can be stored in the Datasheet.

Expression Format

- Addition: *a* **+** *b*
- Subtraction: *a* **-** *b*
- Multiplication: *a* ***** *b*
- Division: *a* **/** *b*
- Power: *a* **^** *b*
- Group: **(** *a* **)**
- Function: *function_name***(** *a* **)**
- Constant *π*: pi

a and *b* can be numbers, column numbers, column variable names (delimited by "") or valid expressions of the format described above. *function_name* has to be a valid function name (described below). The parser is case-sensitive (uppercase and lowercase letters are different). Parentheses must be matched.

Operator Precedence and Associativity

- Parentheses have precedence over all operators.
- The power operator (^) has precedence over +,-,*, and /.
- * and / have precedence over binary operators + and -.
- Unary - has precedence over binary + and -.
- The power operator (^) is right associative, and all other operators are left associative.

Available Functions

Functions must be invoked in the form of *function_name***(** *a* **)**, where the argument *a* is a valid expression.

abs
: absolute value

arccos
: arc cosine / cosine inverse

arcsin
: arc sine / sine inverse

arctan
: arc tangent / tangent inverse

ceil
: ceiling (smallest integer that is not less than the argument)

cos
: cosine

exp
: Euler's number e raised to the power of the argument

factorial
: factorial x! = x(x-1)(x-2)...1

floor
: floor (largest integer that is not greater than the argument)

ln
: natural logarithm

log
: logarithm base 10

round
: round to the closest integer, obtained by floor(*a* + 0.5)

sin
: sine

sqrt
: square root

tan
: tangent

For Trigonometric Functions

Select whether the input is specified in radians or degrees. The conversions from radians to degrees and from degrees to radians are generally inexact. Thus, do not expect arccos(cos(*x*)) to be exactly *x* degrees.

Input Interface

Enter a mathematical expression directly into the expression text box, or use the provided buttons and function list to construct your input expression. The numbers 0 to 9, the decimal point ., and the constant π are available in the **Numbers** panel. The operators +, -, *, /, ^, √ (square root), e^ (exponential), and parentheses are available in the **Operators** panel. The **Functions** panel contains the list of mathematical functions available. To use a function in the list, first click on the function name, then click the **Select** button. For trignometric functions, select radians or degrees mode using the corresponding radio buttons. Click **OK** to evaluate the input expression. The result is displayed the the results panel. Click **Back** for backspace, and **Clear** to clear the input box and results panel.

Store Results in Datasheet (optional)

Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name (e.g. C1) in the provided text box.

Calculate > Probability Distributions

The probability distribution utilities perform the following computations for a number of commonly used probability distributions:

- Probability density (Given *x*, find *P(x)*)
- cumulative probability (Given *x*, find *P(<=x)*)
- Inverse cumulative probability (Given cumulative probability *A*, find *x* such that *P(<=x) = A*)

You can access these utilities from the **Calculate > Probability Distributions** menu. In the probability distribution dialog, you will specify the parameters of the distribution (for example, mean, standard deviation, or degrees of freedom). Input data can be a column of values from a Datasheet, or it can be a single constant. Click the **Compute** button to perform the selected computation. The dialog will remain open until the user clicks the **Close** button.

The following probability distributions are supported:

Continuous

- Normal
- Student's t
- Chi-square
- F
- Uniform

Discrete

- Binomial
- Discrete
- Geometric
- Integer
- Poisson

Calculate > Probability Distributions > Normal

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the normal distribution.

Distribution Parameters:

- Specify the mean of the normal distribution.
- Specify the standard deviation of the distribution.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Student's t

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the Student's t distribution.

Distribution Parameters:

- Specify the degrees of freedom.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Chi-square

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the Chi-square distribution.

Distribution Parameters:

- Specify the degrees of freedom.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > F

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the F (Fisher) distribution.

Distribution Parameters:

- Specify the numerator degrees of freedom.
- Specify the denominator degrees of freedom.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Exponential

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the exponential distribution.

The exponential probability density function, where $\lambda > 0$ is the rate parameter, is defined as:

$$f(x) = \lambda e^{-\lambda x}$$

The cumulative and inverse cumulative probability distribution function is defined as:

$$F(x) = 1 - e^{-\lambda x}$$

$$x = \frac{ln(1 - F(x))}{-\lambda}$$

Distribution Parameters:

- Specify the rate parameter of the distribution.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. To store results in a column, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Uniform

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the uniform distribution.

Distribution Parameters:

- Specify the lower bound of the distribution.
- Specify the upper bound of the distribution.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Binomial

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the binomial distribution.

Distribution Parameters:

- Specify the number of trials (must be a positive integer).
- Specify the event probability (must be a nonnegative number).

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Discrete

This utility computes the probability density, cumulative probability, and inverse cumulative probability of a custom discrete distribution. You can specify the values of a random variable in one column and their corresponding probabilities in another, and use them as inputs to this utility.

Distribution Parameters:

- Select the column containing values of a random variable.
- Select the column containing the corresonding probabilities.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Geometric

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the geometric distribution. You can choose between the two variants of the geometric distribution: (1) the probability distribution of the number of trials needed to get the first success, and (2) the probability distribution of the number of failures before the first success.

Distribution Parameters:

- Specify the success probability (must be between 0 and 1).
- Choose the variant of the geometric distribution:
    - Model the number of trials needed to produce the first success
    - Model the number of failures before the first success

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Integer

Each integer in the integer distribution has the same probability. This utility computes the probability density, cumulative probability, and inverse cumulative probability of the integer distribution. You can specify the minimum and maximum values of the distribution.

Distribution Parameters:

- Specify the minimum of the distribution.
- Specify the maximum of the distribution.

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > Probability Distributions > Poisson

This utility computes the probability density, cumulative probability, and inverse cumulative probability of the Poisson distribution.

Distribution Parameters:

- Specify the mean value (must be a positive number).

Compute:

- Select the desired type of computation.

Input(s):

- If the input data is in a column in the current Datasheet, select the **Column** radio button and the column name in the drop-down menu.
- In the input data is a single constant, select the **Constant** radio button and enter the input value in the provided text box.

Store Results in: (optional)

- Computed results can be stored in a column of the current Datasheet. If storing results is desired, enter the column name or variable name in the provided text box.

Calculate > p-Value

The p-Value tool computes p-Values for four different types of probability distribution (Normal, Student's t, Chi-square, and F) and three types of tests (left-tail, right-tail, two-tail), given a test statistic and degrees of freedom (if applicable).

The inputs to the tool are:

- **Distribution**: Normal, Student's t, Chi-square, or Fisher
- **Type of Test**: Left-tail, Right-tail, or Two-tail
- **Test Statistic**: Sample statistic based on the sample data
- **Degrees of Freedom** (DOF): Number of values that are free to vary after certain restrictions have been imposed on all values. Computation of p-Values for different distributions requires different number of DOF inputs:
  - Normal: none
  - Student's T and Chi-square: one DOF input
  - F: DOF 1 and DOF 2
- Click **Compute** to start the p-Value computation. The result is shown on the log window.
- Click **Clear** to clear all inputs (set to default).

Calculate > Frequency Table

This utility computes the frequencies of data belonging to different categories or classes. You must provide the data (category names or numbers) to be counted in a column. If the data is numerical, you must provide the upper limits of the classes for which frequencies are computed. The categories/classes and their corresponding frequencies are computed and stored in columns specified by the user.

Dialog Inputs

- **Source Data:** Select the column containing input data in the **Compute frequency data for column:** drop-down menu.
- **Methods of Computing Frequencies:** Choose one of the following:
  - Treat source data as categories. Count the frequency of each category.
  - Source data contains numerical data. Count the frequency of each class, where the upper class limits are provided (you must select the column containing the upper class limits in the corresponding drop-down menu).
  - Source data contains numerical data. Count the frequency of each class, where the classes are defined by the first upper class limit and the class width. The number of classes can also be provided; if it is not provided, the number of classes is the smallest number such that all samples can be classified.
  - Source data contains numerical data. Automatically use 10 classes, with the first lower class limit decided by the minimum value. The class width is (maximum value - minimum value) / 10.
- **Store Frequency Table in:** Enter the column name (e.g. C1) or variable name for storing the categories and frequencies.

Statistics

The Statistics menu provides access to descriptive and inferential statistics dialogs.

Statistics > Basic Statistics

This menu provides descriptive statistical computations on rows and columns.

Statistics > Basic Statistics > Descriptive Statistics

Descriptive Statistics computes descriptive statistics for a set of input column variables or for groups of data within a set of input column variables. It displays the results in the log window and also optionally stores them in Datasheets.

To use the utility, use the following steps:

- Select **Statistics > Basic Statistics > Descriptive Statistics** to open the utility.
- In the **Input Variable(s)** text box, enter the names of the columns. Separate column names by spaces (e.g. c2 c4), and use a dash to indicate a range of columns (e.g. c10-c15).
- In the **By Variable** drop-down menu, select the variable that will be used to group data values in each column.
- Select the descriptive statistics to be computed by checking their corresponding check boxes.
- Check the **Store Results in New Datasheet** box if you would like to store the statistics in new Datasheet(s). Statistics for each input variable will be stored in separate Datasheet.
- Click OK to finish. The results are displayed in the log window and stored in new Datasheet(s) if specified.

The following descriptive statistics are available:

- Mean
- SE of mean
- Standard deviation
- Variance
- Coefficient of variation
- First quartile
- Median
- Third quartile
- Interquartile range
- Mode
- Percentile
- Trimmed mean
- Sum
- Minimum
- Maximum
- Range
- Sum of squares
- Skewness
- Kurtosis
- MSSD
- N nonmissing
- N missing
- N total
- Cumulative N
- Percent
- Cumulative percent

Formulas > Descriptive Statistics > Mean

The *mean* (or arithmetic mean) of a data set is the sum of the observations divided by the number of observations. It is a measure of center.

The mean μ of a set *x* of *n* values is

$$\mu = \frac{\sum x}{n}$$

Formulas > Descriptive Statistics > Standard Error of the Mean

The standard error of the mean, or the standard deviation of the sample means, is calculated by dividing the standard deviation by square root of the sample size.

$$SEMean = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation and *n* is the number of nonmissing observations.

Formulas > Descriptive Statistics > Standard deviation

The *standard deviation* of a set of sample values is a measure of variation of values about the mean. It is computed using the following formula:

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

Formulas > Descriptive Statistics > Variance

The **variance** of a set of values is the square of the standard deviation. It is a measure of variation.

Formulas > Descriptive Statistics > Coefficient of Variation

The **coefficient of variation (CV)** is the percentage of the standard deviation relative to the mean.

$$CV = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Formulas > Descriptive Statistics > First Quartile

The **first quartile** (Q1) separates the bottom 25% of the sorted values in a data set from the top 75%.

For a data set $x = \{x_1, x_2, \ldots, x_n\}$, in which the values are sorted in increasing order,

$$Q1 = x_i + a(x_{i+1} - x_i)$$

where
i = floor((n+1)/4)
a = (n+1)/4 - i.

Formulas > Descriptive Statistics > Median

The *median* of a data set is the middle value when the data values are arranged in order of increasing magnitude.

The median is found by the following steps:

- Sort the values in increasing order.
- If the number of values is odd, the median is the number located in the exact middle of the list of sorted values.
- If the number of values is even, the median is the mean of the two middle values.

Formulas > Descriptive Statistics > Third Quartile

The **third quartile** (Q3) separates the bottom 75% of the sorted values of a data set from the top 25%.

For a data set $x = \{x_1, x_2, \ldots, x_n\}$, in which the values are sorted in increasing order,

$$Q3 = x_i + a(x_{i+1} - x_i)$$

where

$$i = \text{floor}(3(n+1)/4)$$
$$a = 3(n+1)/4 - i.$$

Formulas > Descriptive Statistics > Interquartile Range

The **interquartile range** is the difference between the first and third quartiles.

Formulas > Descriptive Statistics > Mode

The **mode** of a data set is the value the occurs most frequently. If no value is repeated, there is no mode. **N for mode** in the results indicates the frequency of the mode.

Formulas > Descriptive Statistics > Percentile

The p^{th} **percentile** separates the bottom p% of the sorted values in a data set from the top p%.

For a data set $x = \{x_1, x_2, \ldots, x_n\}$, in which the values are sorted in ascending order,

$$p^{th} \text{ percentile} = \begin{cases} x_1 & \text{if } i \le 0 \\ x_n & \text{if } i \ge n \\ x_i + a(x_{i+1} - x_i) & \text{otherwise} \end{cases}$$

where

$$i = \text{floor}(p(n+1)/100)$$

$$a = p(n+1)/100 - i.$$

Formulas > Descriptive Statistics > Trimmed Mean

The trimmed mean of a data set is the mean of the data set in which a certain percentage of extreme values are removed. The *x*% trimmed mean is found by arranging the data in order, deleting the top *x*% and the bottom *x*% of the values, and then calculating the mean of the remaining values.

Formulas > Descriptive Statistics > Sum

The *sum* of a set of values is obtained by adding all the values in the data set.

The sum of a set $x = x_1, x_2, ...x_n$ is

$$\sum x = x_1 + x_2 + ... + x_n$$

Formulas > Descriptive Statistics > Minimum

The *minimum* of a data set is the lowest value of the set.

Formula > Descriptive Statistics > Maximum

The *maximum* of a data set refers to the largest value in the set.

Formula > Descriptive Statistics > Range

The *range* of a data set is the difference between the maximum and the minimum.

Formulas > Descriptive Statistics > Sum of Squares

The sum of squares of a data set $x = x_1, x_2, ...x_n$ is the sum of the squares of the data value $\sum x_i^2 = x_1^2 + x_2^2 + ... + x_n^2$.

Formulas > Descriptive Statistics > Skewness

Skewness is a measure of assymetry of a distribution. A positive skew indicates that the right tail is longer and the mass of the distribution is concentrated on the left. A negative skew indicates that the left tail is longer and the mass of the distribution is concentrated on the right.

The formula for unbiased skewness is

$$skewness = \frac{n}{(n-1)(n-2)} \cdot \sum z^3$$

where

$$z = \frac{x - \bar{x}}{\sigma}$$

x = data set
n = number of data values in x
$\bar{x}$ = mean of the data set x
σ = standard deviation of the data set

The formula for biased skewness is

$$\frac{\sqrt{n}\sum(x-\bar{x})^3}{(\sum(x-\bar{x})^2)^{3/2}}$$

Formulas > Descriptive Statistics > Kurtosis

Kurtosis is a measure of the peakedness of a distribution. A high kurtosis indicates a sharper peak and flatter tails, whereas a low kurtosis indicates a rounder peak and wider shoulders.

The formula for kurtosis that is unbiased and centered at 0 is

$$kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \sum z^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where

$$z = \frac{x - \bar{x}}{\sigma}$$

x = data set
n = number of data values in x
$\bar{x}$ = mean of the data set x
σ = standard deviation of the data set

The formula for kurtosis that is biased and centered at 3 is

$$\frac{n \sum (x - \bar{x})^4}{(\sum (x - \bar{x})^2)^2}$$

Formulas > Descriptive Statistics > MSSD

MSSD stands for Mean Squared Successive Differences. It is computed using the following formula:

$$MSSD = \frac{\sum(x_i - x_{i-1})^2}{n-1}$$

where

x_i = the ith data value in the set x
n = the number of data values in set x

Formulas > Descriptive Statistics > N nonmissing

N nonmissing refers to the number of nonmissing observations in a data set.

Formulas > Descriptive Statistics > N missing

N missing refers to the number of missing observations in a data set.

Formulas > Descriptive Statistics > N total

N total is the total number of observations in a data set. *N total* = *N missing* + *N nonmissing*

Formulas > Descriptive Statistics > Cumulative N

Cumulative N is the cumulative frequency of the number of nonmissing observations in a column, or the cumulative frequency for each group in a column, as defined by a BY variable.

Formulas > Descriptive Statistics > Percent

The statistic **percent** expresses the percentage of the frequency of data values in a group out of the total number of nonmissing observations in a column variable.

Formulas > Descriptive Statistics > Cumulative Percent

The statistic **Cumulative Percent** refers to the cumulative percentage of a series of groups. The cumulation is computed in the order of the successive groups presented in the results.

Statistics > Basic Statistics > Column Statistics

The **Column Statistic** utility computes various statistics on the data within a specific column in a Datasheet. The computed statistics are displayed in the log window and can be optionally stored in the Datasheet. You can access this utility through **Calculate > Column Statistics...**. The column statistics of a column can be obtained through the following steps:

- Select **Statistics > Basic Statistics > Column Statistics...**.
- In the **Column Input Variable** drop-down menu, select the column for which you want to calculate statistics.
  - Sum
  - Mean
  - Standard deviation
  - Minimum
  - Maximum
  - Range
  - Median
  - Sum of squares
  - N total
  - N nonmissing
  - N missing
- [optional] If you want to store the computed statistics in a column of the Datasheet, select the **Column** radio button and enter the column number (*e.g.* C1 for column 1) or the variable name of the column (*e.g.* Ages for the variable Ages).
- [optional] If you want to store the computed statistics in a row of the Datasheet, select the **Row** radio button and enter the row number (*e.g.* 3 for row 3).
- Click OK. The results will be shown in the sessioin window and be put in the Datasheet if the store option was used.

Statistics > Basic Statistics > Row Statistics

The **row statistics** utility computes a statistic for each row in a set of columns and stores the results in the corresponding rows of a new column.

The row statistics of a set of input variables can be obtained using the following steps:

- Select **Statistics > Basic Statistics > Row Statistics...**
- In the **Input Variable(s)** list, enter the names of the columns. Separate column names by spaces (e.g. c2 c4), and use a dash to indicate a range of columns (e.g. c10-c15).
- Select the desired statistic:
  - Sum
  - Mean
  - Standard deviation
  - Minimum
  - Maximum
  - Range
  - Median
  - Sum of squares
  - N total
  - N nonmissing
  - N missing
- Enter the column name (e.g. C1) or variable name (e.g. sum) in the **Store Results in:** text box.
- Click OK to finish. The statistic computed across the input column variables for each row will be stored in the corresponding row of the specifed column for storing results.

Statistics > Basic Statistics > Normality Test

The normality test utility performs the Ryan-Joiner normality test and creates a normal quantile plot for a set of sample data values. The Ryan-Joiner normality test computes a correlation coefficient and critical value used to test the claim that a population is normally distributed. The null hypothesis of the test is that the population is normal. The test statistic is the correlation coefficient *r*:

$$\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

where *x* and *y* are the ordered observations and their normal scores, respectively. The normal scores for the ordered observations are the inverse cumulative normal probabilities for *(i+1)/(n+1)*, where *n* is the sample size and *i=0...n-1*.

The critical values for significance level α = 0.10, 0.05, and 0.01 are approximated as follows:

$$CV(n) \approx 1.0071 - \frac{0.1371}{\sqrt{n}} - \frac{0.3682}{n} + \frac{0.7780}{n^2}\ [\alpha = 0.10]$$

$$CV(n) \approx 1.0063 - \frac{0.1288}{\sqrt{n}} - \frac{0.6118}{n} + \frac{1.3505}{n^2}\ [\alpha = 0.05]$$

$$CV(n) \approx 0.9963 - \frac{0.0211}{\sqrt{n}} - \frac{1.4106}{n} + \frac{3.1791}{n^2}\ [\alpha = 0.01]$$

If the test statistic *r* is less than the critical value, reject the null hypothesis that the population is normal; otherwise, fail to reject the null hypothesis.

This utility allows the user to specify a column containing the sample data values. The user can select the significance level used for the Ryan-Joiner normality test. It also provides the option of generating a normal quantile plot (along with the options of having a plot title, axis labels, a regression line, and whether the data values should be used as x or y coordinates).

- To open the utility, select **Statistics > Basic Statistics > Normality Test**.
- Select the column containing the data values in the **Input Variable** drop-down menu.
- Select the **significance level** used in the Ryan-Joiner normality test.
- Select the **Generate normal quantile plot** check box if a normal quantile plot for the input data should be displayed.
    - Enter the plot title in the **Plot Title:** text field.
    - Enter labels for the x and y axis in the corresponding text fields.
    - Select the **Show regression line** check box to show a best-fit line to the points on the plot.
- Click **OK**. The results of the Ryan-Joiner normality test will be displayed in the log window.

Statistics > Sample Size

The **Statistics > Sample Size** menu provides tools for determining the sample size required to estimate population mean and proportion so that a given confidence level and margin of error are met.

Statistics > Sample Size > 1-Population Mean

This tool determines the sample size required to estimate a population mean. You will need to provide the confidence level, desired margin of error, and population standard deviation. The sample size is computed using the following formula:

$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

where $z_{\alpha/2}$ is the inverse cumulative probability of the z distribution at 1 - $\alpha/2$, σ is the population standard deviation, and E is the margin of error.

To use this tool, provide the following inputs to the dialog:

- **Confidence Level**: a number between 0 and 1. For example, enter 0.95 for a 95% confidence level.
- **Standard Deviation**: the population standard deviation.
- **Margin of Error**: the desired margin of error.

Statistics > Sample Size > 1-Population Proportion

This tool determines the sample size required to estimate a population proportion. You will need to provide the confidence level and desired margin of error. You can also provide a proportion estimate if known. If a proportion estimate is given, the sample size is computed using the following formula:

$$n = \frac{\left[z_{\alpha/2}\right]^2 \hat{p}\hat{q}}{E^2}$$

where $z_{\alpha/2}$ is the inverse cumulative probability of the z distribution at 1 - α/2, $\hat{p}$ is the proportion estimate, $\hat{q} = 1 - \hat{p}$, and E is the margin of error.

If a proportion estimate is not given, the following formula is used:

$$n = \frac{\left[z_{\alpha/2}\right]^2 \cdot 0.25}{E^2}$$

To use this tool, provide the following inputs to the dialog:

- **Confidence Level**: a number between 0 and 1. For example, enter 0.95 for a 95% confidence level.
- **Proportion Estimate**: an estimate of the population proportion.
- **Margin of Error**: the desired margin of error.

Confidence Intervals

A confidence interval is an interval of values used to estimate the true value of a population parameter, such mean, proportion, and variance. Statcato provides utilities that compute the confidence intervals for one population and two populations. To access these utilities, go to the **Statistics > Confidence Intervals** menu.

Statistics > Confidence Intervals > 1-Population Mean

This utility computes the confidence intervals for a population mean in one population, whether the population standard deviation is known or unknown. The confidence intervals are calculated using the following formulas:

$$\bar{x} - E < \mu < \bar{x} + E$$

where $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ when population standard deviation σ is known

or $E = t_{\alpha/2}\frac{s}{\sqrt{n}}$ when σ is unknown;

$\bar{x}$ = sample mean

μ = population mean

s = sample standard deviation

$t_{\alpha/2}$ = inverse cumulative probability of the Student's t distribution at 1 - α/2.

$z_{\alpha/2}$ = inverse cumulative probability of the z distribution at 1 - α/2.

To use the utility, select **Statistics > Confidence Intervals > 1-Population Mean**.

- If population standard deviation is known, select the **Known** radio button, and enter the standard deviation in the given text box. Otherwise, select the **Unknown** radio button.
- You can provide sample data in a column of the Datasheet or provide summary data. If your sample data is available in one column, select the **Samples in column:** radio button, and enter the column names (Enter valid column names separated by space. For a continuous range of columns, separate using dash (e.g. C1-C30).).
- If individual sample data is not available, you can provide summary information of the sample data. In this case, select the **Summarized sample data** radio button. Enter the sample size, mean, and standard deviation (only if population standard deviation is unknown) in the provided text boxes.
- Enter the confidence level (between 0 and 1) in the **Confidence level:** text box.
- Click **Ok** to compute the confidence interval. The results will be shown in the log window.

Statistics > Confidence Intervals > 1-Population Proportion

This utility computes the confidence intervals for a population proportion in one population using normal approximation. The confidnece intervals are calculated using the following formula:

$\hat{p} - E < p < \hat{p} + E$ where $E = z_{\alpha/s}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

n = sample size
$\hat{p}$ = sample proportion
$z_{\alpha/2}$ = inverse cumulative probability of the z distribution at 1 - α/2.

If the conditions for normal approximation ($np \geq 5$, $nq \geq 5$) are not met, a warning will be shown along with the results of the computation.

To use the utility, select **Statistics > Confidence Intervals > 1-Population Variance**.

- You can provide sample data in a column of the Datasheet or provide summary data. If you provide sample data, put sample values of at most two categories in a column in the Datasheet. Select the **Samples in column:** radio button, and select the column name in the given drop-down menu.
- If individual sample data is not available, you can provide summary information of the sample data. In this case, select the **Summarized sample data** radio button. Enter the number of trials and the number of events in the provided text boxes.
- Enter the confidence level (between 0 and 1) in the **Confidence level:** text box.
- Click **OK** to compute the confidence interval. The results will be shown in the log window.

The number of trials, number of events, and proportion in the sample are displayed, along with the resulting the margin of error and the confidence interval for the population proportion. Here is an example:

```
Confidence Interval - One population proportion: confidence level = 0.95
Input: Summary data
```

Number of trials	Number of Events	Sample proportion	Margin of Error	95.00%CI
100	10	0.100	0.059	(0.0412, 0.1588)

Statistics > Confidence Intervals > 1-Population Variance

This utility computes the confidence intervals for a population variance in one population. The confidnece intervals are calculated using the following formula:

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$ where

n = sample size
s = sample standard deviation
σ^2 = variance
χ^2_L = inverse cumulative probability of the Chi-square distribution at α/2.
χ^2_R = inverse cumulative probability of the Chi-square distribution at 1 - α/2.

To use the utility, select **Statistics > Confidence Intervals > 1-Population Variance**.

- You can provide sample data in a column of the Datasheet or provide summary data. To provde sample data in a column, select the **Samples in column:** radio button, and select the column name in the given drop-down menu.
- If individual sample data is not available, you can provide summary information of the sample data. In this case, select the **Summarized sample data** radio button. Enter sample size and either sample variance or sample standard deviation in the provided text boxes.
- Enter the confidence level (between 0 and 1) in the **Confidence level:** text box.
- Click **OK** to compute the confidence interval. The results will be shown in the log window.

The sample size, variance, and standard deviation are displayed, along with the resulting confidence interval for variance and standard deviation. Here is an example:

```
Confidence Interval - One population variance: confidence level = 0.95
Input: Summary data
```

N	Variance	Stdev	95.00%CI Variance	95.00%CI Stdev
100	10.000	3.162	(7.7090, 13.4949)	(2.7765, 3.6735)

Statistics > Confidence Intervals > 2-Population Means

This utility computes confidence intervals for the difference of two population means from two independent samples. The confidence interval estimate of μ_1 - μ_2 is:

$$(\bar{x_1} - \bar{x_2}) - E < (\mu_1 - \mu_2) < (\bar{x_1} - \bar{x_2}) + E$$

where E is the margin of error, $\bar{x}$ is the sample mean.

If population standard deviations σ_1 and σ_2 are unknown and are not assumed equal, the margin of error is given by

$$E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where n is the sample size, s is the sample standard deviation, and $t_{\alpha/2}$ is the inverse cumulative probability of Student's t distribution at 1 - α/2. The degree of freedom is given by

$$dof = \frac{(A+B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}}$$

where $A = s_1^2 / n_1$ and $B = s_2^2 / n_2$.

If population standard deviations σ_1 and σ_2 are unknown and are assumed equal, the margin of error is given by

$$E = t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

where $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$, and the degree of freedom is $n_1 + n_2 - 2$.

If population standard deviations σ_1 and σ_2 are known, the margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Dialog Inputs

The sample data of the population must be of only two categories. They can be inputted in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The sample size, mean, and standard deviation of each of the two populations are provided (instead of individual sample values).

If **population standard deviations are known**, check the appropriate check box and provide the standard deviations.

Check the **Assume population variances are equal** check box if the population variances are assumed equal.

The **confidence level** must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

Statistics > Confidence Intervals > 2-Population Proportion

This utility computes confidence intervals for the difference of two population proportions. The confidence interval estimate of p_1 - p_2 is:

$$(\hat{p_1} - \hat{p_2}) - E < (p_1 - p_2) < (\hat{p_1} - \hat{p_2}) + E$$

where the margin of error *E* is given by $E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$, $\hat{p}$ is the sample proportion, and $\hat{q} = 1 - \hat{p}$.

Dialog Inputs

The sample data of the population must be of only two categories. They can be inputted in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The number of events and the number of trials of each of the two populations are provided (instead of individual sample values).

The confidence level must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

Statistics > Confidence Intervals > 2-Population Variance

This utility computes the confidence intervals for the ratio of two population variances (σ_1^2 / σ_2^2). The confidence intervals are calculated using the following formula:

$$\left(\frac{s_1^2}{s_2^2}\frac{1}{F_R}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{s_1^2}{s_2^2}\frac{1}{F_L}\right)$$

where s_1 and s_2 are the sample standard deviations of two population samples, F_L is the lower critical value (inverse F probability distribution at α/2, where α is the significance level), and F_R is the upper critical value (inverse F probability distribution at 1 - α/2).

Dialog Inputs

The sample data can be inputted in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The sample size and standard deviation/variance of each sample are provided (instead of individual sample values).

The confidence level must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

Statistics > Confidence Intervals > Matched Pairs

This utility computes confidence intervals for the mean of the differences between matched pairs. The confidence interval estimate of the mean of the differences between matched paris μ_d is:

$$\bar{d} - E < \mu_d < \bar{d} + E$$

where the margin of error *E* is given by $E = t_{\alpha/2}\frac{s_d}{\sqrt{n}}$, $\bar{d}$ is the mean of the differences between sample matched pairs, and μ_d is the mean of the differences between population matched pairs.

Dialog Inputs

The sample data can be inputted in one of two ways:

- **Samples in columns**: The sample matched pairs are provided in two columns, and the two values in a matched pair are provided in the same row. Select the column containing values of the first sample in **First Sample:** drop-down menu. Select the column containing values of the second sample in **Second Sample:** drop-down menu.
- **Summarized Sample Data**: The sample size, mean, and standard deviation are provided (instead of individual sample values).

The confidence level must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

Statistics > Hypothesis Tests

The **Hypothesis Tests** utilities calculates statistics that test hypotheses about population mean, proportion, and variance.

Statistics > Hypothesis Tests > 1-Population Mean

This utility performs calculations for testing claims about a population mean for the case population standard deviation σ is known and the case when σ is unknown. The population is assumed to be normally distributed. The null hypothesis of a claim about a population mean μ is $\mu = \mu_0$. The alternative hypothesis can be one of the following: $\mu < \mu_0$, $\mu > \mu_0$, or $\mu \neq \mu_0$.

When σ is known, the standard normal distribution is used to calculate p-Values and critical values. The test statistic is computed using the following formula:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where $\bar{x}$ is the sample mean, μ is the hypothesized population mean, and *n* is the sample size.

When σ is unknown, the Student's t distribution with degrees of freedom *n - 1*, where *n* is the sample size, is used to compute p-Values and critical values. The test statistic is computed as follows:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where *s* is the sample standard deviation.

To use the utility, follow these steps:

- **Population standard deviation:** Select the **Known** radio button if the population standard deviation is known and enter the standard deviation in the provided text field. If it is unknown, select the **Unknown** radio button.
- If individual samples are entered in a single column of the Datasheet, select the **Samples in column:** radio button, and select the column name in the drop-down menu.
- To use summary statistics of the sample data, select the **Summarized sample data:** radio button, and input the sample size, mean, and standard deviation in the provided text fields.
- Enter the significance or confidence level (between 0 and 1).
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu. Enter the hypothesized population mean in the provided text box.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Sample Outputs

The null hypothesis and the alternative hypothesis are displayed. The results, along with the input parameters, are displayed in a table.
[N = sample size; Sample Mean = calculated from the individual samples or provided directly by the user; Stdev = s for sample standard deviation, or σ for population standard deviation; Significance Level = 1 - confidence level; Critical Value = critical value corresponding to the significance level; Test Statistic = test statistic corresponding to the hypothesized mean; p-Value = p-Value corresponding to the test statistic]

```
Hypothesis Test - One Population Mean: confidence level = 0.95
Input: C1
σ unknown
Null hypothesis: μ = 98.6
Alternative hypothesis: μ < 98.6
```

N	Sample Mean	Stdev s	Significance Level	Critical Value	Test Statistic	p-Value
12	98.392	0.535	0.05	-1.796	-1.349	0.1023

Hypothesis Test - One Population Mean: confidence level = 0.95
Input: C1
σ known
Assumed population standard deviation σ = 0.62
Null hypothesis: μ = 98.6
Alternative hypothesis: μ ≠ 98.6

N	Sample Mean	Stdev σ	Significance Level	Critical Value	Test Statistic	p-Value
12	98.392	0.620	0.05	1.960	-1.164	0.2444

Statistics > Hypothesis Tests > 1-Population Proportion

This utility performs calculations for testing claims about a population proportion. A normal distribution is used to approximate the binomial distribution. The null hypothesis of a claim about a population proportion p is $p = p_0$. The alternative hypothesis can be one of the following: $p < p_0$, $p > p_0$, or $p \neq p_0$.

The test statistic is computed using the following formula:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where $\hat{p}$ is the sample proportion, p is the hypothesized population proportion, and n is the sample size.

To use the utility, follow these steps:

- If individual samples are entered in a single column of the Datasheet, select the **Samples in column:** radio button, and select the column name in the drop-down menu.
- To use summary statistics of the sample data, select the **Summarized sample data:** radio button, and input the number of events and the number of trials in the provided text fields.
- Enter the significance or confidence level (between 0 and 1).
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu. Enter the hypothesized population proportion in the provided text box.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Sample Outputs

The null hypothesis and the alternative hypothesis are displayed. The results, along with the input parameters, are displayed in a table.
[N = sample size; Sample Proportion = calculated from the individual samples or provided directly by the user; Significance Level = 1 - confidence level; Critical Value = critical value corresponding to the significance level; Test Statistic = test statistic corresponding to the hypothesized proportion; p-Value = p-Value corresponding to the test statistic]

```
Hypothesis Test - One population proportion: confidence level = 0.95
Input: Summary data
Null hypothesis: p = 0.1
Alternative hypothesis: p > 0.1
```

N	Sample Proportion	Significance Level	Critical Value	Test Statistic	p-Value
100	0.200	0.05	1.645	3.333	0.0004

Statistics > Hypothesis Tests > 1-Population Variance

This utility performs calculations for testing claims about a population variance (or standard deviation). The population is assumed to have a normal distribution. The null hypothesis of a claim about a population variance σ^2 is $\sigma = \sigma_0^2$. The alternative hypothesis can be one of the following: $\sigma < \sigma_0$, $\sigma > \sigma_0$, or $\sigma \neq \sigma_0$.

The test statistic is computed using the following formula:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

where *s* is the sample standard deviation, σ is the hypothesized population standard deviation, and *n* is the sample size.

To use the utility, follow these steps:

- If individual samples are entered in a single column of the Datasheet, select the **Samples in column:** radio button, and select the column name in the drop-down menu.
- To use summary statistics of the sample data, select the **Summarized sample data:** radio button, and input the sample size in the provided text field. Select the **Variance** radio button to provide the sample variance, and enter the sample variance in the given text field. To provide the sample standard deviation, select the **Standard deviation** radio button and input it in the corresponding text field.
- Enter the significance or confidence level (between 0 and 1).
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu. Enter the hypothesized population variance in the provided text box.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Sample Outputs

The null hypothesis and the alternative hypothesis are displayed. The results, along with the input parameters, are displayed in a table.
[N = sample size; Sample Stdev s = sample standard deviation calculated from the individual samples or provided directly by the user; Sample Var s2 = sample variance; Significance Level = 1 - confidence level; Critical Value = critical value corresponding to the significance level; Test Statistic = test statistic corresponding to the hypothesized variance; p-Value = p-Value corresponding to the test statistic]

```
Hypothesis Test - One population variance: confidence level = 0.95
Input: Summary data
Null hypothesis: σ2 = 1.44
Alternative hypothesis: σ2 ≠ 1.44
```

N	Sample Stdev s	Sample Var s2	Significance Level	Critical Value	Test Statistic	p-Value
100	1.500	2.250	0.05	73.361, 128.422	154.688	0.0006

Statistics > Hypothesis Tests > 2-Population Means

This utility performs calculations for testing claims about the difference of two population means from two independent samples. The null hypothesis of a claim about the difference of two population means μ_1 *and* μ_2 is $\mu_1 = \mu_2$. The alternative hypothesis can be one of the following: $\mu_1 < \mu_2$, $\mu_1 > \mu_2$, or $\mu_1 \neq \mu_2$.

If population standard deviations σ_1 and σ_2 are unknown and are assumed to be unequal, the test statistics t is given by:

$$t = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} - \frac{s_2^2}{n_2}}}$$

where $\bar{x}$ is the sample mean, μ is the hypothesized population mean, s is the sample standard deviation, and n is the sample size. The degree of freedom is given by

$$dof = \frac{(A+B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}}$$

where $A = s_1^2 / n_1$ and $B = s_2^2 / n_2$.

If population standard deviations σ_1 and σ_2 are unknown and are assumed to be equal, the test statistics t is given by:

$$t = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} - \frac{s_p^2}{n_2}}}$$

where $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$.

If population standard deviations σ_1 and σ_2 are known, the test statistics z is given by:

$$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$$

The null and alternative hypotheses, degrees of freedom (DOF), critical value(s), test statistic, and p-Value are displayed in the outputs.

Dialog Inputs

The sample data can be provided in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The sample size, mean, and standard deviation of each of the two populations are provided (instead of individual sample values).

If **population standard deviations are known**, check the appropriate check box and provide the standard deviations.

Check the **Assume population variances are equal** check box if the population variances are assumed equal.

The **Significance Level** or **Confidence Level** must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

The **Alternative Hypothesis** can be one of the following forms: **Less than** ($\mu_1 < \mu_2$), **Not Equal to** ($\mu_1 \neq \mu_2$), or **Greater than** ($\mu_1 > \mu_2$). Enter the **Hypothesized Mean** ($\mu_1 - \mu_2$) in the provided text box.

Statistics > Hypothesis Tests > 2-Population Proportions

This utility performs calculations for testing claims about the difference of two population proportions. A normal distribution is used for approximation. The null hypothesis of a claim about the difference of two population proportions p_1 *and* p_2 is $p_1 = p_2$. The alternative hypothesis can be one of the following: $p_1 < p_2$, $p_1 > p_2$, or $p_1 \neq p_2$.

The user has the option of using a pooled estimate of the sample proportions $p_1 = x_1 / n_1$ *and* $p_2 = x_2 / n_2$. The pooled esimate is given by:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

The test statistics z, calculated using the pooled estimate is:

$$z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\bar{p_1}(1-\bar{p})}{n_1} + \frac{\bar{p_2}(1-\bar{p_2})}{n_2}}}$$

where $\hat{p}$ is the sample proportion, p is the hypothesized population proportion, and n is the sample size.

The test statistics calculated without using the pooled estimate is given by:

$$z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\hat{p_1}(1-\hat{p})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}}$$

Dialog Inputs

The sample data of the population must be of only two categories. They can be provided in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The number of events and the number of trials of each of the two populations are provided (instead of individual sample values).

The **Significance Level** or **Confidence Level** must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

The **Alternative Hypothesis** can be one of the following forms: **Less than** ($p_1 < p_2$), **Not Equal to** ($p_1 \neq p_2$), or **Greater than** ($p_1 > p_2$). Enter the **Hypothesized Proportion Difference** (p_1 - p_2) in the provided text box.

Select the **Use pooled estimate** check box if the pooled estimate is appropriate for combining the samples.

Statistics > Hypothesis Tests > 2-Population Variance

This utility performs calculations for testing claims about the two population variances. The null hypothesis of a claim about the two population proportions σ_1^2 *and* σ_2^2 is $\sigma_1^2 = \sigma_2^2$. The alternative hypothesis can be one of the following: $\sigma_1^2 < \sigma_2^2$, $\sigma_1^2 > \sigma_2^2$, or $\sigma_1^2 \neq \sigma_2^2$.

The test statistic F is equal to σ_1^2 / σ_2^2.

The test statistic F, p-Value, and critical values of the F distribution given a user-specified significance level are displayed as results.

Dialog Inputs

The sample data can be inputted in one of three ways:

- **Samples in one column**: The population labels of samples are in one column of the Datasheet, and the individual samples are in another column.
- **Samples in two columns**: The samples of the two population are in two separate columns.
- **Summarized sample data:** The sample size and standard deviation/variance of each sample are provided (instead of individual sample values).

The significance level or confidence level must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

The **Alternative Hypothesis** can be one of the following forms: **Less than** ($\sigma_1^2 < \sigma_2^2$), **Not Equal to** ($\sigma_1^2 \neq \sigma_2^2$), or **Greater than** ($\sigma_1^2 > \sigma_2^2$).

Statistics > Hypothesis Test > Matched Pairs

This utility performs calculations for testing claims about the mean of the differences between a population of matched pairs. The population is assumed to be approximately normal. The null hypothesis of a claim about a population mean μ is $\mu = \mu_d$. The alternative hypothesis can be one of the following: $\mu < \mu_d$, $\mu > \mu_d$, or $\mu \neq \mu_d$.

The hypothesis test statistic is given by:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

where $\bar{d}$ is the mean of the differences between sample matched pairs, s_d is the standard deviation of the differences between sample matched pairs, and n is the number of matched pairs. The degrees of freedom is $n - 1$.

Dialog Inputs

The sample data can be inputted in one of two ways:

- **Samples in columns**: The sample matched pairs are provided in two columns, and the two values in a matched pair are provided in the same row. Select the column containing values of the first sample in **First Sample:** drop-down menu. Select the column containing values of the second sample in **Second Sample:** drop-down menu.
- **Summarized Sample Data**: The sample size, mean, and standard deviation are provided (instead of individual sample values).

The significance level or confidence level must be between 0 and 1. For example, enter 0.95 for a 95% confidence level.

The **Alternative Hypothesis** can be one of the following forms: **Less than** ($\mu < \mu_d$), **Not Equal to** ($\mu \neq \mu_d$), or **Greater than** ($\mu > \mu_d$). Enter the **Hypothesized Mean Difference** (hypothesized mean of differences for the population of all matched pairs) in the provided text box.

Statistics > Hypothesis Tests > Hypothesis Test Conclusion Tool

The Hypothesis Test Conclusion Tool makes the initial conclusion of whether to reject the null hypothesis of a hypothesis test and gives the wording of the final conclusion of the test. To use this tool, you give the original claim in symbolic form (e.g. mean ≤ 3, proportion ≠ 0.5) and provide the appropriate values for either the traditional or the p-Value approaches. Based on the original claim, this tool determines the null and alternative hypotheses and the type of test (left-tail, right-tail, or two-tail). The conclusion will be based on either the critical value and test statistic (traditional method) or the significance level and p-Value (p-Value method).

Original Claim

- Select the name of the population parameter in the **Parameter** drop-down menu. You may also enter a customized name in the drop-down box.
- Select the comparison operator (<, ≤, >, ≥, =, ≠) in the **Condition** drop-down menu.
- Enter the hypothesized value in the **Value** text field.
- Click the **Next** button for the next step.

Hypotheses and Sample Statistics

- The original claim, null hypothesis, and alternative hypothesis are displayed in this section. A diagram indicating the type of test is also displayed.
- Select the method for determining conclusion:
  - **Traditional Method**: Enter the critical value and test statistic. If the test statistic falls in the critical region (as determined by the critical value), reject the null hypothesis; otherwise, fail to reject the null hypothesis.
  - **p-Value Method**: Enter the significance level and p-Value. If the p-Value is less than or equal to the significance level α, reject the null hypothesis; otherwise, fail to reject the null hypothesis.
- Click the **Next** button for the next step.

Conclusion

- The initial conclusion of whether the null hypothesis is rejected is shown.
- The final conclusion about the original claim is also displayed.

Statistics > Correlation and Regression

The utilities in this menu perform computations for correlation and regression for both paired data and multiple variables.

Statistics > Correlation and Regression > Linear Two Variables

This utility performs computations that determine the correlation and regression between two variables. To use this utility, you must provide the paired x (independent/predictor) and y (dependent/response) values in separate columns. You must also provide the significance level for the hypothesis test for correlation.

For correlation, the following values are computed:

- **Linear correlation coefficient**: $r = \frac{n\sum(xy) - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$
- **Test statistic**: $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$
- **Critical values** $r = \sqrt{\frac{t^2}{n-2+t^2}}$, where *t* is the inverse probability distribution of (1 - significance / 2) in the Student's t distribution with *n - 2* degress of freedom
- **p-Value** of the test statistic *r*

For regression equation $Y = b_0 + b_1 x$, the y-intercept b_0 and the slope b_1 are calculated as follows:

$$b_0 = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

The null hypothesis H_0 for a hypothesis test for linear correlation is that there is no linear correlation (ρ = 0), and the alternative hypothesis H_1 is that there is a linear correlation (ρ ≠ 0).

For variation, the following values are computed ($\bar{y}$ is the mean of the *y* variable values, $\hat{y}$ is the *y* value predicted by the regression equation):

- **Explained variation**: $\sum(\hat{y} - \bar{y})^2$
- **Unexplained variation**: $\sum(y - \hat{y})^2$
- **Total variation**: $\sum(y - \bar{y})^2$
- **Coefficient of determination**: $r^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$
- **Standard error of estimate**: $s = \sqrt{\frac{\sum(y - \hat{y})^2}{n-2}}$

Dialog Inputs

- In the **Independent/dependent variable series** list, select the columns containing data values for the input variables. The x variable is the independent variable, and the y variable is the dependent variable. Click **Clear Input List** to clear the input columns list.
- **Significance level**: Enter the significance level of the hypothesis test for correlation.
- Select the **Show a scatterplot for all pairs of data values** check box to display a scatter plot

showing data values in each pair of input variables. Use the provided options to customize the scatterplot.

Statistics > Correlation and Regression > Multiple Regression

This utility performs computations that determine the linear regression between multiple variables. To use this utility, you must provide the values for at least one independent/predictor variables and a dependent/response in separate columns.

The multiple regression equation is

$$\boldsymbol{y} = b_0 + b_1\boldsymbol{x_1} + b_2\boldsymbol{x_2} + \ldots + b_k\boldsymbol{x_k}$$

where $\boldsymbol{y}$ is the independent variable with *n* sample values and $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_k}$ are the dependent variables, each of which also has *n* sample values. The regression coefficients $b_0, b_1, \ldots, b_k$ are determined by least squares estimation.

For variation, the following values are computed ($\bar{y}$ is the mean of the *y* variable values, $\hat{y}$ is the *y* value predicted by the regression equation):

- **Explained variation**: $\sum(\hat{y} - \bar{y})^2$
- **Unexplained variation**: $\sum(y - \hat{y})^2$
- **Total variation**: $\sum(y - \bar{y})^2$
- **Coefficient of determination**: $r^2 = \dfrac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$
- **Adjusted coefficient of determination**: $1 - \dfrac{n-1}{n-(k+1)}(1-r^2)$
- **Standard error of estimate**: $s = \sqrt{\dfrac{\sum(y - \hat{y})^2}{n-k-1}}$
- **Test statistics F**: $F = \dfrac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \hat{y})^2} \cdot \dfrac{n-k-1}{k}$
- **p-Value** corresponding to the test statistics (P(x > the test statistics in the F distribution with numerator degrees of freedom *k* and denominator degrees of freedom *n-k-1*))

Dialog Inputs

Independent Variables

Select an independent variable under the **Select the column containing an independent variable** list. Click the **Add to list** button to add the column to the list of independent variables. To remove a variable from the list, select the variable name and click the **Remove from list** button. Ctrl-click or shift-click to select multiple columns. Click the **Clear Input List** button to clear the input list.

Dependent Variable

Select the dependent variable under the **Select the column containing the dependent variable** drop-down menu.

Click OK to perform the computation. The results will be displayed in the log window.

Statistics > Correlation and Regression > Non-Linear Models

This utility performs computations that fit a non-linear mathematical regression model to two variables. To use this utility, you must provide the values for an independent/predictor variable and a dependent/response variable in separate columns.

The following types of non-linear models are available:

- Quadratic: $y = ax^2 + bx + c$
- Cubic: $y = ax^3 + bx^2 + cx + d$
- Polynomial: $y = ax^n + bx^{n-1} + cx^{n-2} + ...$, where *n* is a positive integer
- Logarithmic: y = a ln(x) + b
- Power: $y = ax^b$
- Fixed Power: $y = ax^n$, where *n* is a real number
- Exponential: $y = ab^x$

where *y* is the independent variable and *x* is the dependent variable.

The non-linear regression problem is solved using least squares estimation. The problem is expressed as an inconsistent system, **A***x* = *b*, where *x* contains regression coefficients, *b* contains values of the dependent variables, and **A** contains values computed using the independent variables such that the chosen non-linear model is represented equivalently by the inconsistent system.

For variation, the following values are computed ($\bar{y}$ is the mean of the *y* variable values, *n* is the sample size, and *k* is the number of regression coefficients):

- **Explained variation**: $SSR = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{b} - n\bar{y}^2$
- **Unexplained variation**: $SSE = \boldsymbol{b}^T \boldsymbol{b} - \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{b}$
- **Total variation**: $SST = \boldsymbol{b}^T \boldsymbol{b} - n\bar{y}^2$
- **Coefficient of determination**: $R^2 = \frac{SSR}{SST}$
- **Standard error of estimate**: $s = \frac{SSE}{n-k-1}$
- **Test statistics F**: $F = \frac{SSR}{SSE} \times \frac{n-k-1}{k}$
- **p-Value** corresponding to the test statistics

Dialog Inputs

- **Select x (independent/predictor variable):** Select the column containing the x values
- **Select y (dependent/response variable):** Select the column containing the y values (the number of x and y values must be the same)
- **Type of Model**: Select desired type.

Statistics > Nonparametrics > Spearman's Rank Correlation

This utility performs computations for the Spearman's rank correlation test, which is a nonparametric test that uses ranks of two samples. To use this utility, you must provide the sample paired data in two separate columns. You must also provide the significance level for the hypothesis test for correlation.

The ranks of the data values are computed for each of the two samples. Let *x* be the ranks of sample 1 and *y* be the ranks of sample 2. The rank correlation coefficient is computed as follows:

$$r_s = \frac{n\sum xy - \sum x \sum y}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

The critical value is

$$r = \frac{t}{\sqrt{n - 2 + t^2}}$$

where *t* is the inverse probability distribution of (1 - significance / 2) in the Student's t distribution with *n - 2* degress of freedom.

The p-value of the test statistic is obtained using a Student's t distribution with *n - 2* degrees of freedom.

Dialog Inputs

- Under the heading **Select the two variables of a new series**, select the two columns containing data values for the input variables. Click the **Add Series** button to add the series to the **Variable series** list. To remove a series from the list, click the **Remove Series** button. Click **Clear Input List** to clear the input columns list.
- **Significance level**: Enter the significance level of the hypothesis test for correlation.

Statistics > Multinomial

The utilities under this menu implement the inferential methods applicable to data that can be separated in different categories.

Statistics > Multinomial > Chi-Square Goodness-of-Fit

The Chi-Square Goodness-of-Fit test determines whether an observed frequency distribution agrees with some claimed distribution. It applies to data of different categories resulting from a multinomial experiment. The test statistic for Goodness-of-Fit tests in multinomial experiments is

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

When using this utility, you must provide the observed frequencies (frequency counts or categorical data) and the expected frequencies or probabilities. You must also provide the confidence level of the test.

Dialog Inputs

Observed Frequencies

- If the frequency counts of the different categories are in one column of the Datasheet, select the **Frequencies in Column:** radio button and select the column in the drop-down menu. If the category names corresponding to the frequency counts are also in one column of the Datasheet, you can optionally select the column containing the category names in the drop-down menu.
- If the categorical data are in one column (instead of frequency counts), selet the **Categorical Data in Column:** radio button and select the column in the drop-down menu.

Expected Frequencies

- If the different categories of data are expected to have the same frequency, select the **Equal Frequencies** radio button.
- If unequal frequencies are expected, select the **Unequal Frequencies** radio button. If the expected frequencies are provided, select the column containing the expected frequencies in the **Frequencies in Column** drop-down menu. If the expected probabilities are given, select the column containing the probabilities in the **Probabilities in Column** drop-down menu.
- If a past sample of categorical data is available, select the **Categorical Data** radio button. Select the column containing the categorical data values in the **Past Sample Data in Column** drop-down menu.

Significance Level

Enter the significance level in the given text box. The significance level must be between 0 and 1.

Statistics > Multinomial > Chi-Square Contingency Table

A contingency table contains frequencies corresponding to two variables, one used to categorize rows and another used to categorize columns. A test of independence tests whether the row and column variables in a contingency table are independent. The null hypothesis of such a test is that the row and column are independent. The test statistic for a test of independence in a contingency table is

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where the expected frequency of a cell in the contingency table is *(row total)(column total)/(grand total)*. The Chi-Square distribution is used, and the degrees of freedom is given by *(number of rows - 1)(number of columns - 1)*.

When using this utility, you must provide the observed frequencies and the confidence level of the test.

Dialog Inputs

Contingency Table

- The columns of a contingency table should be placed in separate columns in the Datasheet. All the columns in a contingency table must have the same number of rows.
- To specify a column of a contingency table, select the column in the **Select a column of the table:** list, and click the **Add to list** button. The columns will be added to the **Columns:** list. Ctrl-click or shift-click to select multiple columns.
- To remove a column that has been added to the list, select the column in the list and click the **Remove from list** button. Click **Clear Input List** to clear the input columns list.

Significance Level

Enter the significance level in the given text box. The significance level must be between 0 and 1.

Statistics > Multinomial > Cross Tabulation and Chi-Square

This utility performs cross tabulation and shows the joint distribution of two variables given individual observations and optionally the frequency of each observation. The joint distribution is displayed in a contingency table, which can be stored in a datasheet for further processing. The option to perform a Chi-square test on the contingency table is also provided.

To use this utility, you will need to provide the observations in two columns (the values of the first variable in one column, and the corresponding values of the second variable in a second column). You may optionally provide the frequency corresponding to each observation in a third column.

Dialog Inputs

- Select the columns containing row and column labels of the observations in the provided drop-down menus.
- If frequencies of the observations are available, select the column containing the frequencies in the provided drop-down menu.
- To perform a Chi-square test, select the coresponding checkbox and provide the significance level of the test.
- To store the resulting contingency table in a new datasheet, select the corresponding checkbox.

Statistics > Analysis of Variance

Analysis of variance (ANOVA) tests the equality of three or more population means by analyzing sample variances. It compares two different estimates of the variance common to the different populations. The utilities in this menu perform one-way ANOVA, where only one characteristic (also referred to as factor or treatment) is used to categorize the populations, and two-way ANOVA, where data are categorized into groups according to two factors.

Statistics > Analysis of Variance > One-Way ANOVA

One-way analysis of variance uses one factor to categorize three or more populations. It is used to test hypotheses that three or more population means are equal. The null hypothesis H_0 is $\mu_1 = \mu_2 = ... = \mu_k$, where μ_i is the mean of population *i* of *k* populations.

The test statistic F is the ratio of two estimates of the variance common to the *k* populations:

$$F = \frac{variance\,between\,samples}{variance\,within\,samples} = \frac{\frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1}}{\frac{\sum (n_i-1)s_i^2}{\sum (n_i-1)}}$$

When the null hypothesis is true, it has an *F* distribution with degrees of freedom given by

numerator degrees of freedom = $k - 1$

denominator degrees of freedom = $N - k$

The test statistics *F* can be expressed as the ratio MS(treatment) / MS (error). The mean square (MS) and sum of squares (SS) are computed in order to obtain *F*.

SS(treatment) is a sum of squares representing the variation between the sample means:

$$SS(treatment) = \sum n_i(\bar{x}_i - \bar{\bar{x}})^2$$

where n_i is the size of sample *i*, $\bar{x}_i$ is the mean of sample *i*, $\bar{\bar{x}} = \frac{\sum n_i \bar{x}_i}{N}$, and *N* is the sum of all sample sizes.

SS(error) is a sum of squares representing the variation that is assumed to be common to the populations:

$$SS(error) = \sum (n_i - 1)s_i^2$$

SS(total) is a measure of the total variation in all the sample data:

SS(total) = SS(treatment) = SS(error)

MS(treatment) is the mean square for treatment:

$$MS(treatment) = \frac{SS(treatment)}{k-1}$$

MS(error) is the mean square for error:

$$MS(error) = \frac{SS(error)}{N-k}$$

To use this utility, you will need to provide sample values for different populations in separate columns and specify the significance level. The SS and MS values are displayed along with the test statistic, critical value, and p-Value.

Dialog Inputs

To include a column in the analysis, select the column name in the **Select the column to be included in the analysis:** list and click the **Add to list** button. The selected column(s) will be added to the **Responses** list. To remove a column from the **Responses** list, select it and click the **Remove from list** button. Ctrl-click or shift-click to select multiple columns.

Click the **Clear Input List** button to clear the input list.

Enter the signicance level (a number between 0 and 1) in the **Significance level** text box.

Statistics > Analysis of Variance > Two-Way ANOVA

Two-way analysis of variance uses two factors to categorize data. It is used to test whether the factors have an effect and whether there is an interaction between the two factors.

Suppose we have the row factor *A* with *a* levels, the column factor *B* with *b* levels, and each treatment group of has *n* independent observations *y*.

		Factor B			
		1	2	...	*b*
Factor A	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$	...	$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$	...	$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$
	...	...	...	...	...
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$	...	$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

Then the row sum is $A_i = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$, the column sum is $B_i = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$, and the cell sum is $AB_{ij} = \sum_{k=1}^{n} y_{ijk}$. The total number of observations (sample size) is *abn*. The number of observations for each level of *A* is *bn*, and the number of observations for each level of *B* is *an*.

The two-way ANOVA computes the sum of squares (SS), mean of squares (MS), degrees of freedom (DOF), F scores and p-Value for the two factors and the interaction between the two factors. SS, DOF, and MS are also computed for the within-group error.

Source	DOF	SS	MS	F	p-Value
A	*a* - 1	$SS(A) = \frac{\sum_{i=1}^{a} A_i^2}{bn} - CM$	SS(A) / DOF(A)	MS(A) / MS (E)	1 - CumulativeProbability(F, DOF(A), DOF(E))
B	*b* - 1	$SS(B) = \frac{\sum_{j=1}^{b} B_j^2}{an} - CM$	SS(B) / DOF(B)	MS(B) / MS(E)	1 - CumulativeProbability(F, DOF(B), DOF(E))
AB (interaction)	(*a* - 1)(*b* - 1)	$SS(AB) = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} (AB)_{ij}^2}{n} - CM - SS(A) - SS(B)$	SS(AB) / DOF(AB)	MS(AB) / MS(E)	1 - CumulativeProbability(F, DOF(AB), DOF(E))
E (error)	*ab*(*n* - 1)	$SS(E) = SS(Total) - SS(A) - SS(B) - SS(AB)$	SS(E) / DOF(E)		
Total	*abn* - 1	$SS(Total) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - CM$			

CM (correction of the mean) is computed as $CM = \frac{(\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk})^2}{abn}$.

To use this utility, you will need to provide all the observations in one column and their corresponding row factor and column factor labels in corresponding cells of two separate columns.

Dialog Inputs

Response

In the given drop-down menu, select the column contains the observations.

Row Factor

Select the column containing the row factor labels corresponding to the observations.

Column Factor

Select the column containing the column factor labels corresponding to the observations.

Statistics > Nonparametrics > 1-Sample Sign Test

This utility performs calculations for testing claims about a population median for the case when the type of population distribution cannot be assumed. The null hypothesis H_0 of a claim is median = $median_0$, where $median_0$ is the hypothesized median. The alternative hypothesis H_1 can be one of the following: median < $median_0$, median > $median_0$, or median ≠ $median_0$.

A sample value above $median_0$ is assigned a positive (+) sign. A sample value below $median_0$ is assigned a negative (-) sign. The test statistic is computed as follows:

- Left-tailed (H_1: median < $median_0$): s = the number of positive (+) signs
- Right-tailed (H_1: median > $median_0$): s = the number of negative (-) signs
- Two-tailed (H_1: median ≠ $median_0$): s = the minimum of the number of (+) signs and the number of (-) signs

The p-value for a one-tailed test (left-tailed or right-tailed) is $P(X \leq s)$ where X is a random variable representing the number of positive signs. The p-value for a two-tailed test is $2P(X \leq s)$.

If the sample size n (excluding values that are equal to the hypothesized mean) is not more than 50, exact binomial probability calculations are used to compute the p-value. If the sample size is greater than 50, the normal distribution is used as an approximation to binomial:

$$P(X \leq s) = P(z < \frac{s + 0.5 - n/2}{\sqrt{n}/2})$$

To use the utility, follow these steps:

- If individual samples are entered in a single column of the Datasheet, select the **Samples in column:** radio button, and select the column name in the drop-down menu.
- To use summary statistics of the sample data, select the **Summarized sample data:** radio button, and input the number of positive signs and the number of negative signs in the provided text fields.
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu. Enter the hypothesized population median in the provided text box.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Statistics > Nonparametrics > 2-Sample Matched-Pair Sign Test

This utility performs calculations for testing claims about two population means for the case when the type of population distribution cannot be assumed, except that the population of the pairwise differences is symmetric. The null hypothesis H_0 of a claim is $\mu_1 = \mu_2$. The alternative hypothesis H_1 can be one of the following: $\mu_1 < \mu_2$, $\mu_1 > \mu_2$, or $\mu_1 \neq \mu_2$.

A positive difference is assigned a positive (+) sign. A negative difference is assigned a negative (-) sign. The test statistic is computed as follows:

- Left-tailed (H_1: $\mu_1 < \mu_2$): s = the number of positive (+) signs
- Right-tailed (H_1: $\mu_1 > \mu_2$): s = the number of negative (-) signs
- Two-tailed (H_1: $\mu_1 \neq \mu_2$): s = the minimum of the number of (+) signs and the number of (-) signs

The p-value for a one-tailed test (left-tailed or right-tailed) is $P(X \le s)$ where X is a random variable representing the number of positive signs. The-value for a two-tailed test is $2P(X \le s)$.

If the sample size n (excluding values that are equal to the hypothesized mean) is not more than 50, exact binomial probability calculations are used to compute the p-value. If the sample size is greater than 50, the normal distribution is used as an approximation to binomial:

$$P(X \le s) = P(z < \frac{s + 0.5 - n/2}{\sqrt{n}/2})$$

To use the utility, follow these steps:

- If individual samples are entered in two columns of the Datasheet, select the **Samples in column:** radio button, and select the column names in the drop-down menu.
- To use summary statistics of the sample data, select the **Summarized sample data:** radio button, and input the number of positive signs and the number of negative signs in the provided text fields.
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Statistics > Nonparametrics > Wilcoxon Signed Rank Test

This utility performs calculations for the Wilcoxon Signed Rank Test, which is a non-parametric hypothesis test for the median of a single sample or of two samples of matched pairs. The null hypothesis H_0 of a claim is median = $median_0$, where $median_0$ is the hypothesized median. The alternative hypothesis H_1 can be one of the following: median < $median_0$, median > $median_0$, or median ≠ $median_0$.

Let *x* be a value in a single sample or the difference of matched pairs in two samples. For each value of *x*, the absolute value of the difference between *x* and the hypothesized median $median_0$ is computed. Differences of 0 are discarded. Let *n* be the number of nonzero differences. Ranks from 1 to *n* are then assigned to each *x* based on the ascending order of the differences. Mean of ranks are assigned to tied values. A negative (-) sign is assigned to a rank if its corresponding *x* is below $median_0$, and a positive (+) sign is assigned if *x* is above $median_0$.

The rank sums are calculated as follows:

- Left-tailed (H_1: median < $median_0$): $R_{(+)}$ = the sum of ranks with positive (+) signs
- Right-tailed (H_1: median > $median_0$): $R_{(-)}$ = the sum of ranks with negative (-) signs
- Two-tailed (H_1: median ≠ $median_0$): $R_{(+-)}$ = the minimum of $R_{(+)}$ and $R_{(-)}$

If *n* is less than or equal to 30, the exact p-value and critical value for the rank sum is calculated. The test statistic *R* is the rank sum corresponding to the type of test described above. The p-value for a one-tailed test (left-tailed or right-tailed) is $P(X \le R)$ where *X* is a random variable representing the rank sum. The p-value for a two-tailed test is $2P(X \le s)$.

If *n* is greater than 30, normal approximation is used. The test statistic is

$$z_w = \frac{R - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

The p-value for a one-tailed test (left-tailed or right-tailed) is $P(z \le z_w)$ where *z* is a random variable representing the z-score. The p-value for a two-tailed test is $2P(z \le z_w)$.

To use the utility, follow these steps:

- If one sample of values are provided in a single column, select the **One Sample** radio button and the column name in the corresponding drop-down menu.
- If matched pairs of two samples are provided in two columns, select the **Two Samples (Matched Pairs)** radio button, and select the column names in the corresponding drop-down menus.
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu. Enter the hypothesized population median in the provided text box.
- Provide the significance level of the test (between 0 and 1).
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Statistics > Nonparametrics > Wilcoxon Rank Sum Test

This utility performs calculations for the Wilcoxon Rank Sum Test (also known as Mann-Whitney), which is a non-parametric test that uses ranks of samples from two independent populations to test whether two populations have the same distribution. The null hypothesis H_0 of a claim is $median_1 = median_2$. The alternative hypothesis H_1 can be one of the following: $median_1 < median_2$, $median_1 > median_2$, or $median_1 \neq median_2$.

Let n_1 be the size of sample 1 and n_2 be the size of sample 2. The two samples are combined into one and then ranked, with the smallest sample given rank 1, second rank 2, etc. If two or more samples are tied, their ranks are averaged.

The rank sums are calculated as follows:

- R_1 = sum of ranks for sample 1
- R_2 = sum or ranks for sample 2
- $R = R_1$

Let μ_R and σ_R be the mean and standard deviation or the sample R values that is expected when the two populations have the same distribution, respectively:

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Using normal approximation, the test statistic is

$$z_R = \frac{R - \mu_R}{\sigma_R}$$

A continuity correction of +0.5 is used for a left-tailed test, and -0.5 is used for a right-tailed and two-tailed test.

The p-value, where z is a random variable representing the z-score is computed as follows:

- Left-tailed test: $P(z \leq z_R)$
- Right-tailed test: $P(z \geq z_R)$
- Two-tailed test: $2P(z \geq z_R)$ if $z_R > 0$ or $2P(z \leq z_R)$ if $z_R < 0$
  .

To use the utility, follow these steps:

- For inputs, select the names of the two sample columns in the corresponding drop-down menus.
- Select the form of the alternative hypothesis in the **Alternative Hypothesis:** drop-down menu.
- Provide the significance level of the test (between 0 and 1).
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Statistics > Nonparametrics > Kruskal-Wallis Test

This utility performs calculations for the Krustal-Wallis Test, which is a non-parametric test that uses ranks of sample data from three or more independent populations to test the null hypothesis that the independent samples come from populations with the same distribution. The null hypothesis H_0 of a claim is that the samples come from populations with the same distribution (the population medians are all equal). The alternative hypothesis H_1 is that the samples come from populations with different distributions (the population medians are not all equal).

Let k be the number of samples, and N be the total number of observations in all k samples combined. Let n_i be the size of sample i. The k samples are combined into one and then ranked, with the smallest sample given rank 1, second rank 2, etc. If two or more samples are tied, their ranks are averaged. Let R_i be the sum of ranks for sample i.

The test statistic is

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

The test uses an approximation by a Chi-square distribution with $k - 1$ degrees of freedom and computes the right-tailed p-value and critical value.

To use the utility, you must provide at least three samples, each with at least five observations (if a sample of fewer than five observations is used, a note indicating the presence of small samples will be shown).

- To include a column in the analysis, select the column name is the **Select the column to be included in the analysis:** list and click the **Add to list** button. The column will be added to the **Responses** list. To remove a column from the **Responses** list, select it and click the **Remove from list** button.
- Click the **Clear Input List** button to clear the input list.
- Provide the significance level of the test (between 0 and 1).
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Statistics > Nonparametrics > Runs Test

This utility performs calculations for the runs test, which is a non-parametric test that uses the number of runs in a sequence of sample data to test for randomness in the order of the data. A run is a sequence of data belonging to the same category. The null hypothesis H_0 of a claim is that the sequence is random. The alternative hypothesis H_1 is that the sequence is not random.

Each data value can be categorized into one of two separate categories. For example, the two categories can be text descriptions "a" and "b". For numerical data, the two categories can be defined to be above and below a certain numerical value, such as the mean.

Let n_1 be the number of samples belonging to the first category. Let n_2 be the number of samples belonging to the second category. Let G be the number of runs.

Using normal approximation, the test statistic z is

$$z = \frac{G - \mu}{\sigma}$$

where

$$\mu = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

The two-tailed p-value and critical value are provided in the test results.

To use the utility, you must provide the sample data in a single column or provide the summary data. If a sample of fewer than ten observations is used, a note indicating the presence of small samples will be shown.

- If data samples are provided in a column, select the **Samples in one column** radio button and select the column name in the coresponding drop-down menu.
- If summarized data is provided, select the **Summarized sample data** radio button and provide the number of category 1 elements, the number of category 2 elements, and the number of runs.
- Provide the significance level of the test (between 0 and 1).
- If data samples are provided, the **Data Options** panel will be shown. Select the **Categorical** radio button if the samples contain exactly two categories. If the samples contain numerical values, select the **Numerical** radio button and then select the appropriate radio button to categorize the numerical data into above and below the mean or a specified value.
- Click the **OK** button to perform the computation. The results will be displayed in the log window.

Graph

Statcato provides utilies to create different types of graphs for visualizing data.

Once a graph is created, you can access the following functions on the Graph menu:

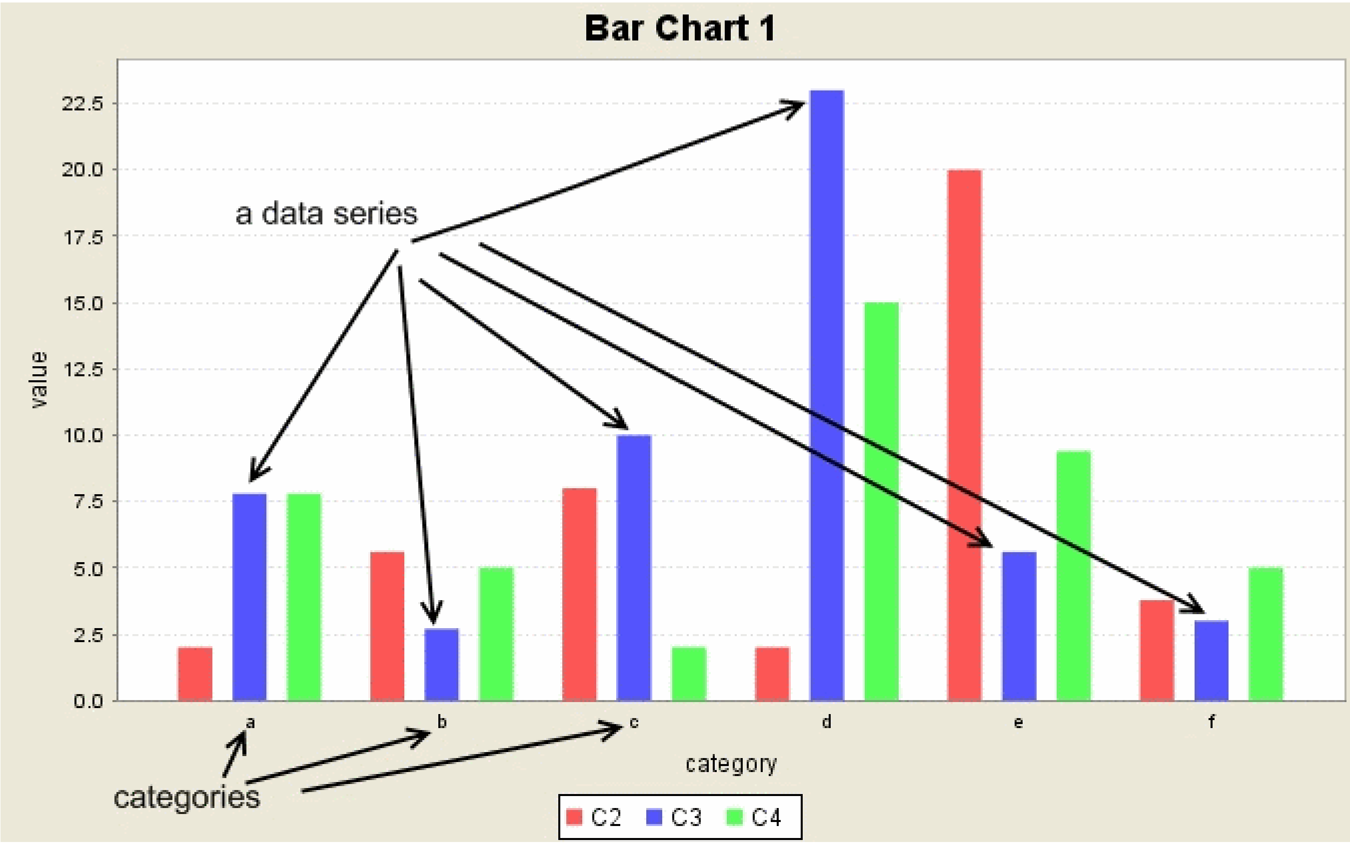
- **Copy Graph to Clipboard**: Copy the graph to clipboard.
- **Save as...**: Save the graph in PNG image format.
- **Print...**: Send the graph to a printer.
- **Properties...**: Change various properties of the graph, such as plot title, axis labels, and background colors.

You can also access the following functions by right-clicking the graph:

- **Properties...**: Change various properties of the graph, such as plot title, axis labels, and background colors.
- **Save as...**: Save the graph in PNG image format.
- **Print...**: Send the graph to a printer.
- **Zoom in** or **Zoom out**
- **Auto range**: Display the graph in a default axis setting.

Graph > Bar Chart

A bar chart shows the frequencies of a number of categories. With this utility, you can create horizontal or vertical bar charts that represent the frequences of a number of data series with data in a number of categories.

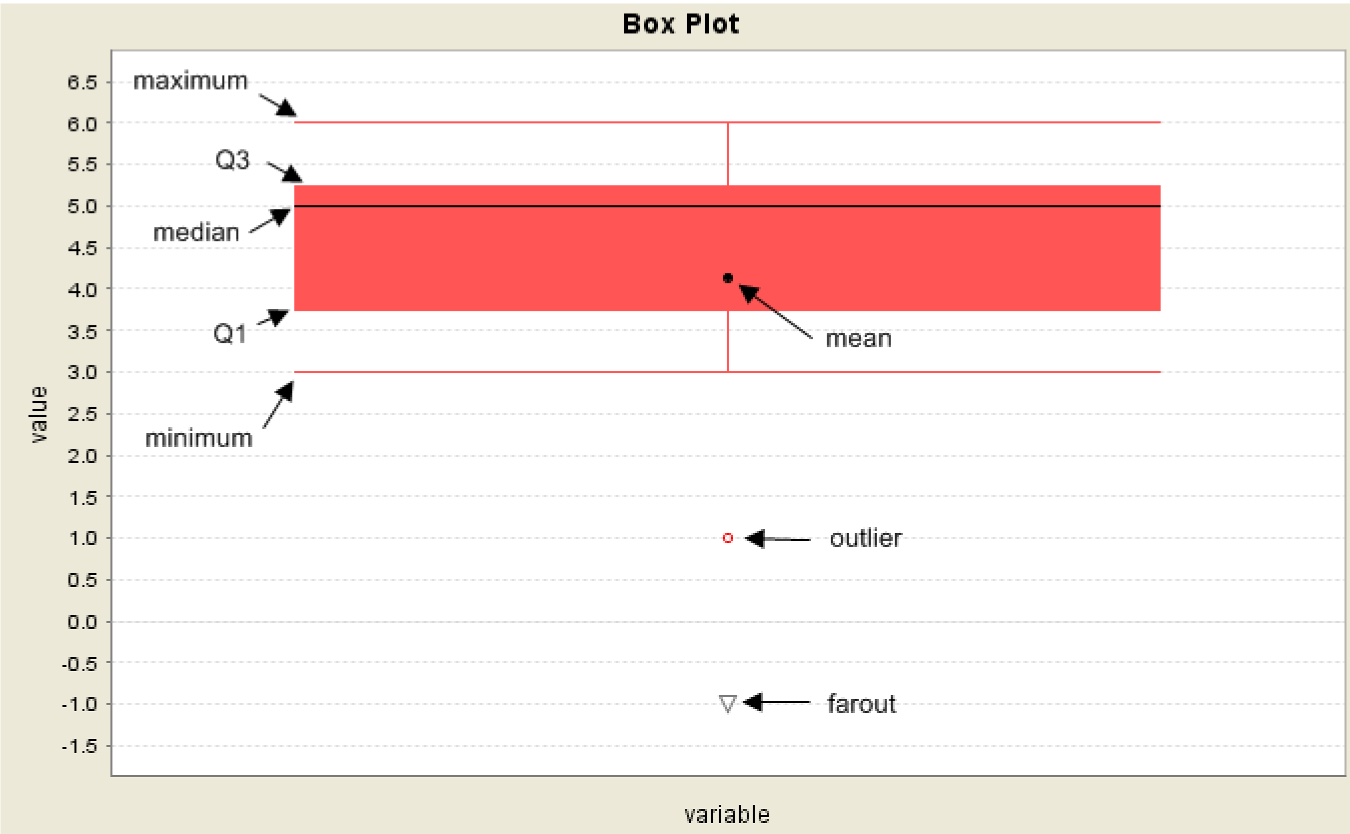


The bar chart utility allows the user to specify graph variables containing data series and a category variable. The number of categories must be the same as the number of data values in each series. It also provides the options of having a plot title, axis labels, and legend.

- To open the histogram utility, select **Graph > Bar Chart**.
- To add a data series, select the column containing the data values in the **Select the column variable of a new series:** and click the **Add Series** button.
- To remove a data series, select the series in the **Graph Series** list, and click the **Remove Series** button.
- Select the column containing category labels for the data values in each series.
- Select whether to have horizontal or vertical bars under **Directions of bars** drop-down menu.
- Enter labels for the x and y axis in the corresponding text fields.
- Enter the plot title in the **Plot Title:** text field.
- Select the **Show Legend** check box to show a legend indicating the different groups.
- Click **OK** to create the bar chart.

Graph > Box Plot

A box plot shows the five-number summaries of a group of data values (minimum, first quartile, median, third quartile, and maximum). The outliers and farout values are shown outside of the box. Outliers are values that are of distances more than 1.5 * IQR (inter-quartile range) but less than or equal to 2 * IQR from the first quartile or the third quartile. Farouts are values that are of distances more than 2 * IQR from the first quartile or the third quartile.



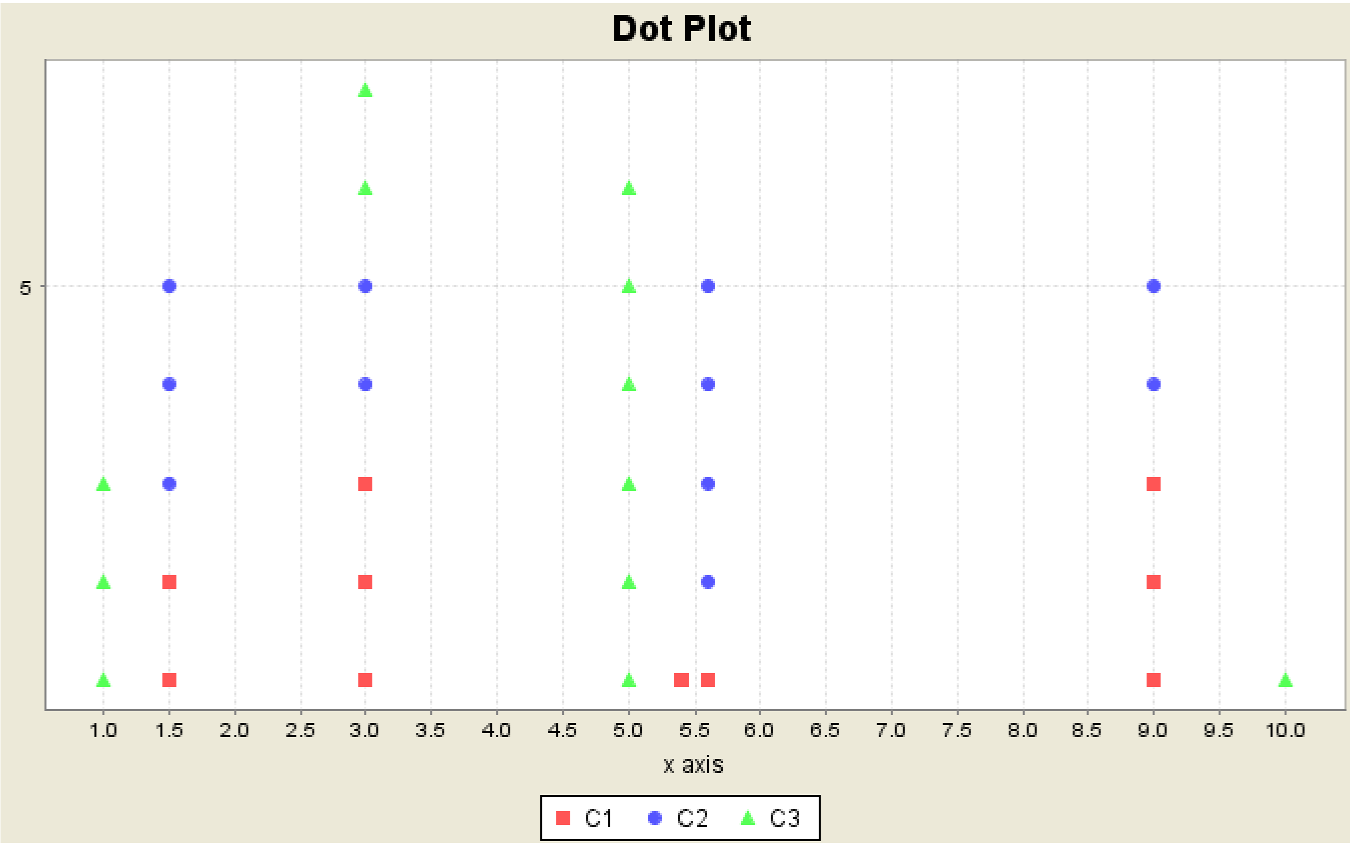
To view the five-number summaries on the box plot generated by Statcato, hover the mouse over the box and the summaries will be displayed as a tooltip.

The box plot graph utility allows the user to specify graph variables containing data values and optionally a group variable. It also provides the options of providing a plot title and showing the legend.

- To open the box plot utility, select **Graph > Box Plot**.
- In the **Graph Variables** list, select the column variable(s) containing data values for which the plot will be created. A separate box is used with each column variable.
- (optional)If the column variables can be categorized by groups, select the column containing the group names corresponding to the data values in the **Grouped By Categories in:** drop-down menu.
- Enter the plot title in the **Plot Title:** text field.
- Enter labels for the x and y axis in the corresponding text fields.
- Select the orientation of the plot (horizontal or vertical).
- Select the **Show Legend** check box to show a legend indicating the different groups.
- Click **OK** to create the box plot.

Graph > Dot Plot

A dot plot shows the occurrences of each distinct data value as a stack of dots. Thus, the height of a stack at a data value represents its frequency.



The box plot graph utility allows the user to specify the graph variable(s) containing data values and a dot plot for them. The dots for each graph variable is shown differently, and the dots for repeated values for different graph variables are stacked.

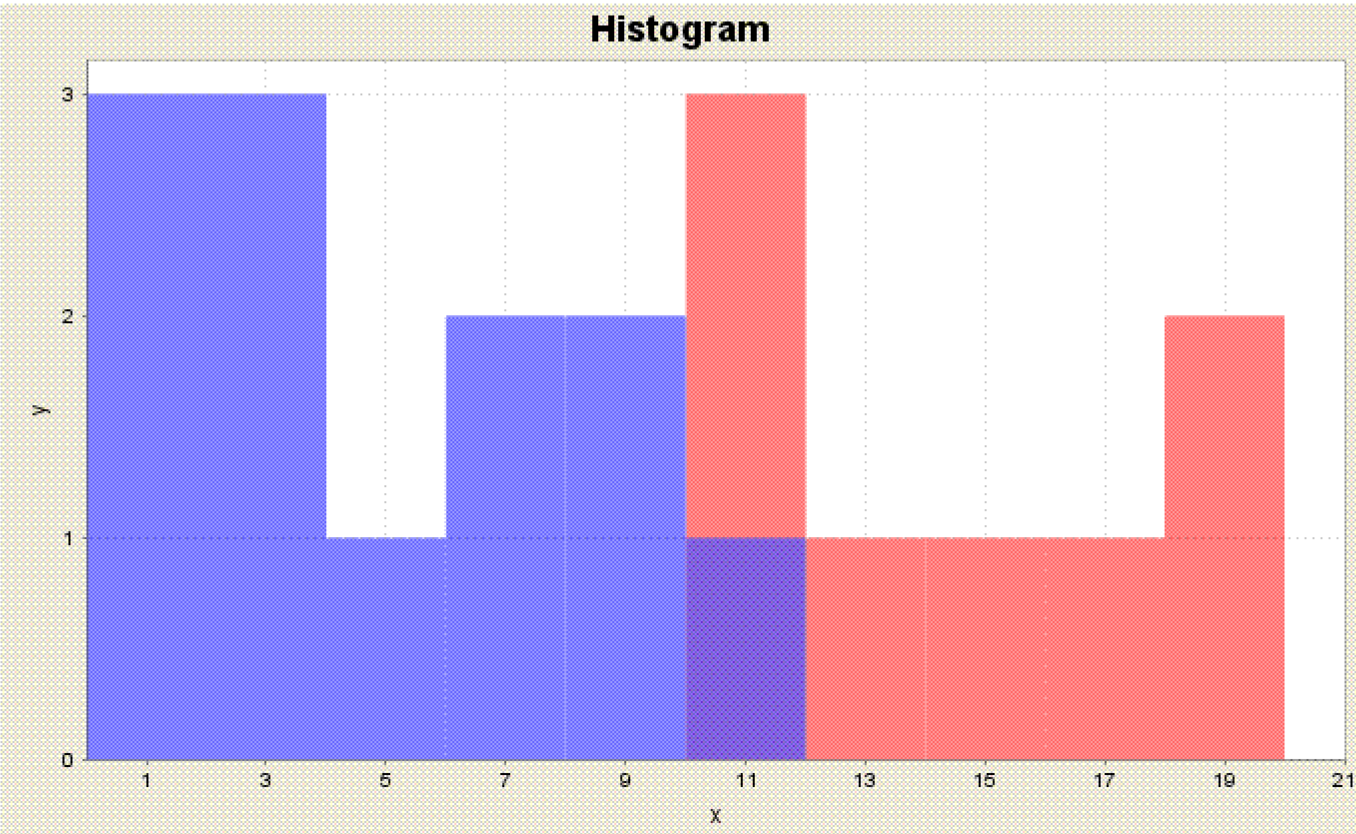
It also provides the options of providing a plot title, specifying the x- and y- axis labels, and showing the legend.

Dialog Inputs

- To open the dot plot utility, select **Graph > Dot Plot**.
- In the **Graph Variables** list, select the column variable(s) containing data values for which the plot will be created. A distinct marker is used with each column variable.
- Enter the plot title in the **Plot Title:** text field.
- Enter the x-axis label and the y-axis label in the corresponding text fields.
- Select the **Show Legend** check box to show a legend indicating the graph variables.
- Click **OK** to create the dot plot.

Graph > Histogram

A histogram shows the frequencies of a number of bins (or classes) with numerical boundaries. The width of each bin is the same. With this utility, you can create histograms that represent the frequences or relative frequencies of a set of bins. Any data value less than the minimum of all values will be assigned to the first bin, and any data value greater than the maximum will be assigned to the last bin. Values at bin boundary will be assigned to the higher bin.



The histogram graph utility allows the user to specify graph variables containing data values and optionally a group variable. It also provides the options of having a plot title, axis labels, and legend.

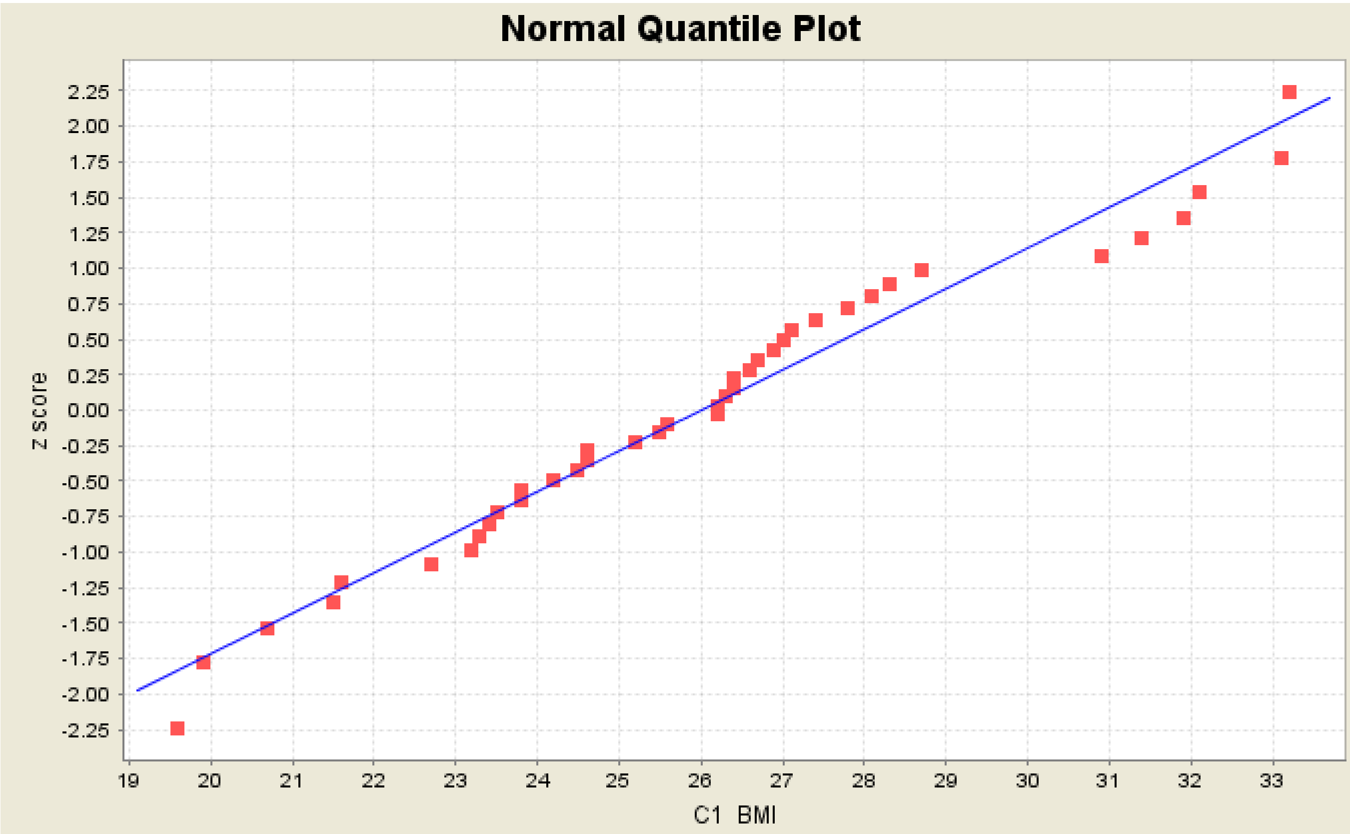
- To open the histogram utility, select **Graph > Histogram**.
- In the **Graph Variables** list, select the column variable(s) containing data values for which the plot will be created. A different color is used for each column variable.
- (optional) If the column variables can be categorized by groups, select the column containing the group names corresponding to the data values in the **Grouped By Categories in:** drop-down menu.
- Select whether to plot the frequencies or the relative frequencies of the bins under **Heights of bars represent:**.
- There are two ways to control binning on the x-axis:
    1. Provide the number of bins/classes, the minimum x value, and the maximum. The class width is determined by (maximum - minimum) / (number of classes). If the minimum and maximum are not provided, they will be the minimum and maximum of all the data values.
    2. Provide the class width and the minimum x value. The number of classes will be the smallest integer such that all data values are classified.
- Other Options:
    - Enter the tick mark units (distance between two consecutive ticks) in the tick mark units text field for the y-axis.
    - Enter the plot title in the **Plot Title:** text field.
    - Enter the labels for the x and y axes.

Select the **Show Legend** check box to show a legend indicating the different groups.

- Click **OK** to create the histogram.

Graph > Normal Quantile Plot

A normal quantile plot is a graph of points (*x, y*) where sample data values are plotted against their expected z scores in a normal distribution corresponding to their positions with respect to other data values. The expected z scores for the sorted sample data values (in ascending order) are the inverse cumulative normal probabilities of *(2i + 1)/(2n)*, where *n* is the sample size and *i=0..n-1*.

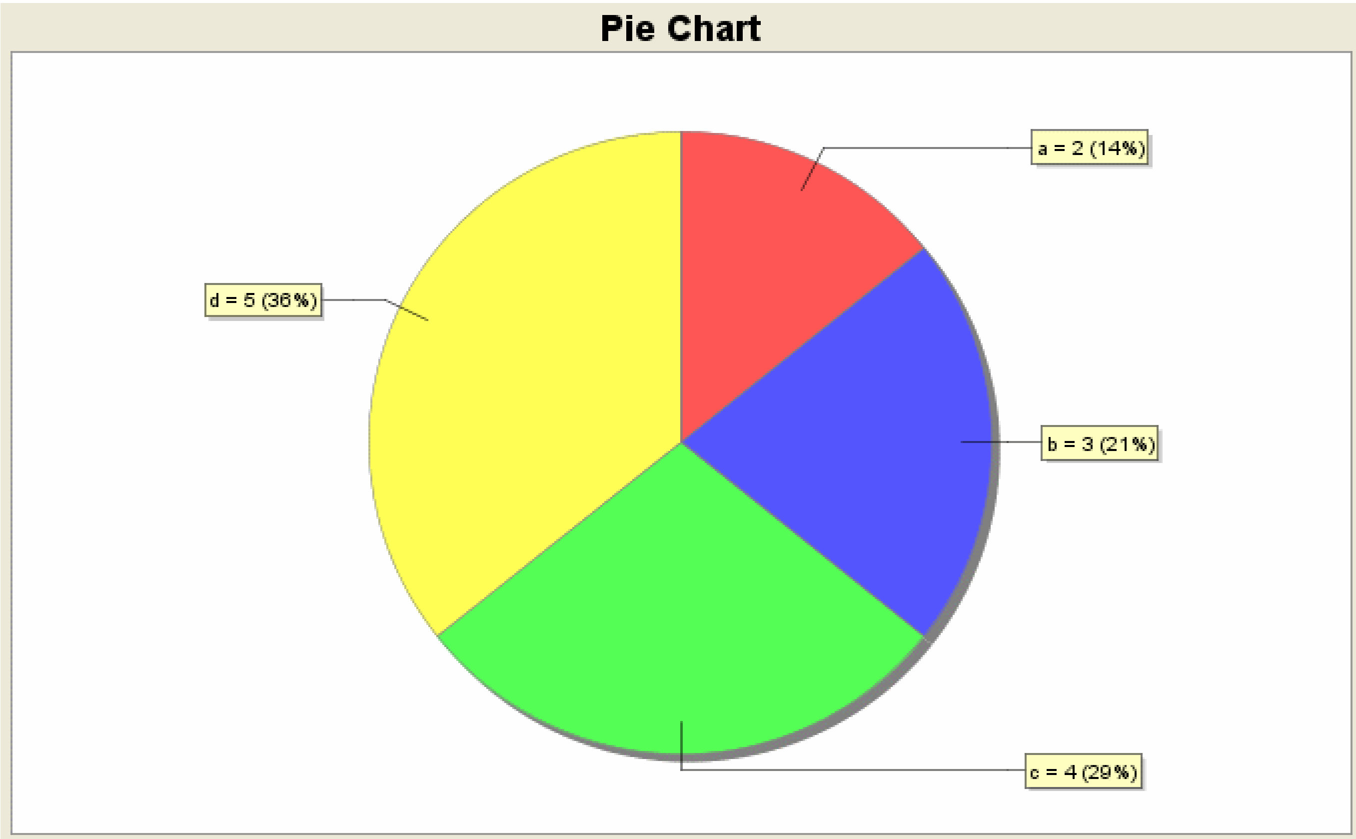


The normal quantile plot utility allows the user to specify a column containing the data values. It also provides the options of having a plot title, axis labels, and a regression line. The user can select whether the data values should be used as x or y coordinates and specify the significance for the Ryan-Joiner normality test.

- To open the plot utility, select **Graph > Normal Quantile Plot**.
- Select the column containing the data values in the **Select the column for which the plot will be created** drop-down menu.
- Enter the plot title in the **Plot Title:** text field.
- Enter labels for the x and y axis in the corresponding text fields.
- Select the **Show regression line** check box to show a best-fit line to the points on the plot.
- Select the significance level used in the Ryan-Joiner normality test.
- Click **OK** to create the plot. The results of the Ryan-Joiner normality test will be displayed in the log window.

Graph > Pie Chart

A pie chart is a circular chart divided into sectors that show relative frequencies of different categories represented.

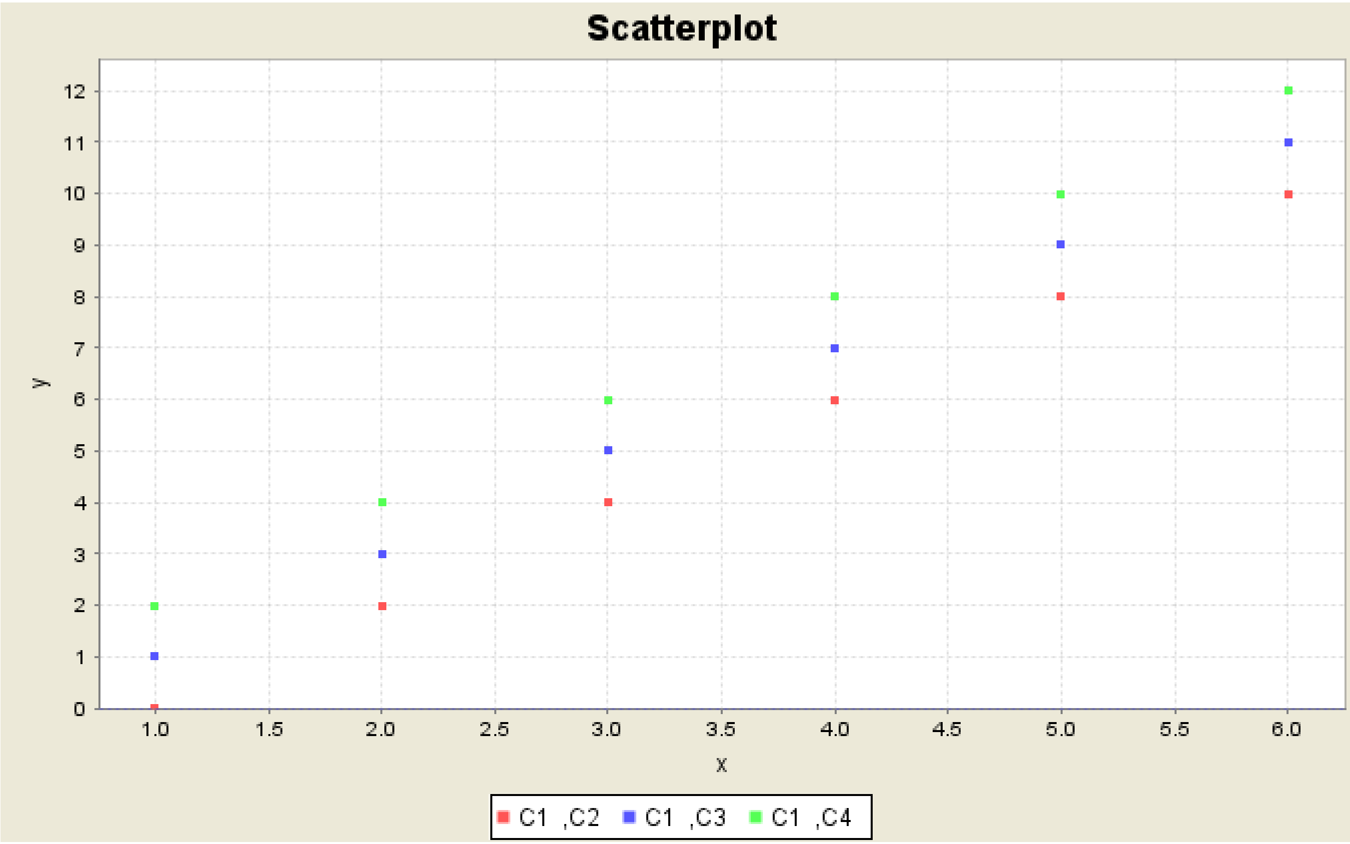


To create a pie chart, use the following steps:

- If summary data of the categories and their corresponding frequencies are present in the Datasheet, select **Summary Data from Datasheet**. Select the column containing categories in the **Categories:** drop-down menu. Select the column containing the respective frequencies in the **Frequencies"** drop-down menu.
- If data values are entered in a single column in the Datasheet, select **Data Values from Datasheet**. In the **Data:** drop-down menu, select the column containing data values.
- Enter the chart title in the **Chart Title:** text field.
- To sort the pie sectors by categories, select **Sort by Categories**. To sort by frequencies, select **Sort by Frequencies**. The sectors are displayed clockwise in an ascending order.
- To display legends, select the **Show Legends** check box.
- To display values/percentages for each pie sector, select **Show Values/Percentages for each Pie Sector**.
- Click **OK** to create the pie chart.

Graph > Scatterplot

A scatterplot shows a collection of points on an x-y plane in cartesian coordinates.



The scatterplot graph utility allows the user to specify a number of series, each of which has a collection of (x, y) points specified by a x column variable and y column variable. It also provides the options of having a plot title, axis labels, legend, and regression lines.

- To open the scatterplot utility, select **Graph > Scatterplot**.
- To add a series of (x, y) points to the plot, select the column variable containing the x coordinates in the **X variable** drop-down menu, select the column variable containing the y coordinates in the **Y variable** drop-down menu, and then click the **Add Series** button. The added series will be shown in the **Graph X,Y Series** list. The x and y column variables must have the same number of items.
- To remove a series from the **Graph X,Y Series** list, select the item and click the **Remove Series** button.
- Enter labels for the x and y axis in the corresponding text fields.
- Enter the plot title in the **Plot Title:** text field.
- Select the **Show Legend** check box to show a legend indicating the different series.
- Select the **Show regression curve** check box to show a regression curve for each data series. Choose the desired regression model, linear or non-linear.
- Click **OK** to create the scatterplot.

Graph > Stem-and-Leaf Plot

A stem-and-leaf plot divides each data value into a stem and a leaf and shows the leaves corresponding to each distinct stem in the data set in a tabular fashion. Like a histogram, the frequency of each category (each stem in a stem-and-leaf plot) can be easily seen. However, unlike a histogram, the individual data value can be retrieved from the plot by combining each stem with its corresponding leaves. For example, for the values {1.5, 2.3, 4.5, 4.8, 6.5, 8.5, 8.8, 9.5}, the stem-and-leaf plot is

Stem	Leaves
1	5
2	3
3	
4	58
5	
6	5
7	
8	58
9	5

In this plot, the unit of each leaf is 0.1. The individual data value can be found by appending the stem to each of its leaf. Stems with no leaves are inserted between the first and the last stem so that all stems are evenly spaced.

The stem-and-leaf graph utility allows the user to specify a number of graph variables and an optional variable that groups the data into a number of categories. It also provides the options of specifying the leaf unit, whether the outliers should be excluded, and whether the stems with no leaves should be displayed. The leaf unit must be equal to 10^x where x is an integer. Outliers are those that fall outside of the interval [first quartile - 1.5 * IQR, third quartile + 1.5 * IQR], where IQR is the interquartile range.

Dialog Inputs

- To open the stem-and-leaf utility, select **Graph > Stem-and-Leaf Plot**.
- Select the columns containing data values under **Graph Variables** list. Control-click to select multiple variables. A separate plot is created for each column of values if no Group-by variable is selected (see next item).
- If the data values have their corresponding category labels in a separate column, select the column of the category labels in the **Group By Categories in** drop-down menu. A separate plot is created for data values belonging to each category label.
- [optional] Enter the **leaf unit** (must be equal to 10^x, where x is an integer). If the leaf unit is not specified, it will be determined by the following method: A leaf unit for a data value is the place to the right of the first non-zero digit of the value (*e.g.*The leaf unit of 141 is 10, and the leaf unit of 0.023 is 0.001). The leaf unit for a set of data values is the maximum of the leaf units of the individual data values.
- Select the **Exclude outliers** checkbox to exclude the outliers from the plot.
- Select the **Display stems that have no leaves** checkbox to display the stems with no leaves.
- Click **OK** to create the stem-and-leaf plot. The plot will be displayed in the log window.

Help > Check for Updates

This tool checks if a new version of Statcato is available. If a new version is available, the **Download the Latest Version** button will be shown. Click to the button to download the latest version to your computer.

References

- Guth, Scott. *Elementary Statistics: Managing Variability and Error.* 1.2 ed. Mt. San Antonio College, 2006.
- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge University Press, 1992.
- Triola, Mario F. *Elementary Statistics*. 10th ed. Pearson Education, Inc., 2006.

GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program--to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS

0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

a) The work must carry prominent notices stating that you modified it, and giving a relevant date.

b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".

c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.

d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.

b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.

c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.

d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.

e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or

b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or

c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or

d) Limiting the use for publicity purposes of names of licensors or authors of the material; or

e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or

f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY

OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.>
Copyright (C) <year> <name of author>

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

<program> Copyright (C) <year> <name of author>
This program comes with ABSOLUTELY NO WARRANTY; for details type `show w'.
This is free software, and you are welcome to redistribute it
under certain conditions; type `show c' for details.

The hypothetical commands `show w' and `show c' should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lgpl.html>.

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

[This is the first released version of the Lesser GPL. It also counts as the successor of the GNU Library Public License, version 2, hence the version number 2.1.]

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public Licenses are intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users.

This license, the Lesser General Public License, applies to some specially designated software packages--typically libraries--of the Free Software Foundation and other authors who decide to use it. You can use it too, but we suggest you first think carefully about whether this license or the ordinary General Public License is the better strategy to use in any particular case, based on the explanations below.

When we speak of free software, we are referring to freedom of use, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish); that you receive source code or can get it if you want it; that you can change the software and use pieces of it in new free programs; and that you are informed that you can do these things.

To protect your rights, we need to make restrictions that forbid distributors to deny you these rights or to ask you to surrender these rights. These restrictions translate to certain responsibilities for you if you distribute copies of the library or if you modify it.

For example, if you distribute copies of the library, whether gratis or for a fee, you must give the recipients all the rights that we gave you. You must make sure that they, too, receive or can get the source code. If you link other code with the library, you must provide complete object files to the recipients, so that they can relink them with the library after making changes to the library and recompiling it. And you must show them these terms so they know their rights.

We protect your rights with a two-step method: (1) we copyright the library, and (2) we offer you this license, which gives you legal permission to copy, distribute and/or modify the library.

To protect each distributor, we want to make it very clear that there is no warranty for the free library. Also, if the library is modified by someone else and passed on, the recipients should know that what they have is not the original version, so that the original author's reputation will not be affected by problems that might be introduced by others.

Finally, software patents pose a constant threat to the existence of any free program. We wish to make sure that a company cannot effectively restrict the users of a free program by obtaining a restrictive license from a patent holder. Therefore, we insist that any patent license obtained for a version of the library must be consistent with the full freedom of use specified in this license.

Most GNU software, including some libraries, is covered by the ordinary GNU General Public License. This license, the GNU Lesser General Public License, applies to certain designated libraries, and is quite different from the ordinary General Public License. We use this license for certain libraries in order to permit linking those libraries into non-free programs.

When a program is linked with a library, whether statically or using a shared library, the combination of the two is legally speaking a combined work, a derivative of the original library. The ordinary General Public License therefore permits such linking only if the entire combination fits its criteria of freedom. The Lesser General Public License permits more lax criteria for linking other code with the library.

We call this license the "Lesser" General Public License because it does Less to protect the user's freedom than the ordinary General Public License. It also provides other free software developers Less of an advantage over competing non-free programs. These disadvantages are the reason we use the ordinary General Public License for many libraries. However, the Lesser license provides advantages in certain special circumstances.

For example, on rare occasions, there may be a special need to encourage the widest possible use of a certain library, so that it becomes a de-facto standard. To achieve this, non-free programs must be allowed to use the library. A more frequent case is that a free library does the same job as widely used non-free libraries. In this case, there is little to gain by limiting the free library to free software only, so we use the Lesser General Public License.

In other cases, permission to use a particular library in non-free programs enables a greater number of people to use a large body of free software. For example, permission to use the GNU C Library in non-free programs enables many more people to use the whole GNU operating system, as well as its variant, the GNU/Linux operating system.

Although the Lesser General Public License is Less protective of the users' freedom, it does ensure that the user of a program that is linked with the Library has the freedom and the wherewithal to run that program using a modified version of the Library.

The precise terms and conditions for copying, distribution and modification follow. Pay close attention to the difference between a "work based on the library" and a "work that uses the library". The former contains code derived from the library, whereas the latter must be combined with the library in order to run.

GNU LESSER GENERAL PUBLIC LICENSE
TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License Agreement applies to any software library or other program which contains a notice placed by the copyright holder or other authorized party saying it may be distributed under the terms of this Lesser General Public License (also called "this License"). Each licensee is addressed as "you".

A "library" means a collection of software functions and/or data prepared so as to be conveniently linked with application programs (which use some of those functions and data) to form executables.

The "Library", below, refers to any such software library or work which has been distributed under these terms. A "work based on the Library" means either the Library or any derivative work under copyright law: that is to say, a work containing the Library or a portion of it, either verbatim or with modifications and/or translated straightforwardly into another language. (Hereinafter, translation is included without limitation in the term "modification".)

"Source code" for a work means the preferred form of the work for making modifications to it. For a library, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the library.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running a program using the Library is not restricted, and output from such a program is covered only if its contents constitute a work based on the Library (independent of the use of the Library in a tool for writing it). Whether that is true depends on what the Library does and what the program that uses the Library does.

1. You may copy and distribute verbatim copies of the Library's complete source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and distribute a copy of this License along with the Library.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Library or any portion of it, thus forming a work based on the Library, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) The modified work must itself be a software library.

b) You must cause the files modified to carry prominent notices stating that you changed the files and the date of any change.

c) You must cause the whole of the work to be licensed at no charge to all third parties under the terms of this License.

d) If a facility in the modified Library refers to a function or a table of data to be supplied by an application program that uses the facility, other than as an argument passed when the facility is invoked, then you must make a good faith effort to ensure that, in the event an application does not supply such function or table, the facility still operates, and performs whatever part of its purpose remains meaningful.

(For example, a function in a library to compute square roots has a purpose that is entirely well-defined independent of the application. Therefore, Subsection 2d requires that any application-supplied function or table used by this function must be optional: if the application does not supply it, the square root function must still compute square roots.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Library, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Library, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Library.

In addition, mere aggregation of another work not based on the Library with the Library (or with a work based on the Library) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may opt to apply the terms of the ordinary GNU General Public License instead of this License to a given copy of the Library. To do this, you must alter all the notices that refer to this License, so that they refer to the ordinary GNU General Public License, version 2, instead of to this License. (If a newer version than version 2 of the

ordinary GNU General Public License has appeared, then you can specify that version instead if you wish.) Do not make any other change in these notices.

Once this change is made in a given copy, it is irreversible for that copy, so the ordinary GNU General Public License applies to all subsequent copies and derivative works made from that copy.

This option is useful when you wish to copy part of the code of the Library into a program that is not a library.

4. You may copy and distribute the Library (or a portion or derivative of it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange.

If distribution of object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place satisfies the requirement to distribute the source code, even though third parties are not compelled to copy the source along with the object code.

5. A program that contains no derivative of any portion of the Library, but is designed to work with the Library by being compiled or linked with it, is called a "work that uses the Library". Such a work, in isolation, is not a derivative work of the Library, and therefore falls outside the scope of this License.

However, linking a "work that uses the Library" with the Library creates an executable that is a derivative of the Library (because it contains portions of the Library), rather than a "work that uses the library". The executable is therefore covered by this License. Section 6 states terms for distribution of such executables.

When a "work that uses the Library" uses material from a header file that is part of the Library, the object code for the work may be a derivative work of the Library even though the source code is not. Whether this is true is especially significant if the work can be linked without the Library, or if the work is itself a library. The threshold for this to be true is not precisely defined by law.

If such an object file uses only numerical parameters, data structure layouts and accessors, and small macros and small inline functions (ten lines or less in length), then the use of the object file is unrestricted, regardless of whether it is legally a derivative work. (Executables containing this object code plus portions of the Library will still fall under Section 6.)

Otherwise, if the work is a derivative of the Library, you may distribute the object code for the work under the terms of Section 6. Any executables containing that work also fall under Section 6, whether or not they are linked directly with the Library itself.

6. As an exception to the Sections above, you may also combine or link a "work that uses the Library" with the Library to produce a work containing portions of the Library, and distribute that work under terms of your choice, provided that the terms permit modification of the work for the customer's own use and reverse engineering for debugging such modifications.

You must give prominent notice with each copy of the work that the Library is used in it and that the Library and its use are covered by this License. You must supply a copy of this License. If the work during execution displays copyright notices, you must include the copyright notice for the Library among them, as well as a reference directing the user to the copy of this License. Also, you must do one of these things:

a) Accompany the work with the complete corresponding machine-readable source code for the Library including whatever changes were used in the work (which must be distributed under Sections 1 and 2 above); and, if the work is an executable linked with the Library, with the complete machine-readable "work that uses the Library", as object code and/or source code, so that the user can modify the Library and then relink to produce a modified executable containing the modified Library. (It is understood that the user who changes the contents of definitions files in the Library will not necessarily be able to recompile the application to use the modified definitions.)

b) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (1) uses at run time a copy of the library already present on the user's computer system, rather than copying library functions into the executable, and (2) will operate properly with a modified version of the library, if the user installs one, as long as the modified version is interface-compatible with the version that the work was made with.

c) Accompany the work with a written offer, valid for at least three years, to give the same user the materials specified in Subsection 6a, above, for a charge no more than the cost of performing this distribution.

d) If distribution of the work is made by offering access to copy from a designated place, offer equivalent access to copy the above specified materials from the same place.

e) Verify that the user has already received a copy of these materials or that you have already sent this user a copy.

For an executable, the required form of the "work that uses the Library" must include any data and utility programs needed for reproducing the executable from it. However, as a special exception, the materials to be distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

It may happen that this requirement contradicts the license restrictions of other proprietary libraries that do not normally accompany the operating system. Such a contradiction means you cannot use both them and the Library together in an executable that you distribute.

7. You may place library facilities that are a work based on the Library side-by-side in a single library together with other library facilities not covered by this License, and distribute such a combined library, provided that the separate distribution of the work based on the Library and of the other library facilities is otherwise permitted, and provided that you do these two things:

a) Accompany the combined library with a copy of the same work based on the Library, uncombined with any other library facilities. This must be distributed under the terms of the Sections above.

b) Give prominent notice with the combined library of the fact that part of it is a work based on the Library, and explaining where to find the accompanying uncombined form of the same work.

8. You may not copy, modify, sublicense, link with, or distribute the Library except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, link with, or distribute the Library is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

9. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Library or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Library (or any work based on the Library), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Library or works based on it.

10. Each time you redistribute the Library (or any work based on the Library), the recipient automatically receives a license from the original licensor to copy, distribute, link with or modify the Library subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties with this License.

11. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not

excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Library at all. For example, if a patent license would not permit royalty-free redistribution of the Library by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Library.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply, and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

12. If the distribution and/or use of the Library is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Library under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

13. The Free Software Foundation may publish revised and/or new versions of the Lesser General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Library specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Library does not specify a license version number, you may choose any version ever published by the Free Software Foundation.

14. If you wish to incorporate parts of the Library into other free programs whose distribution conditions are incompatible with these, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status

of all derivatives of our free software and of promoting the sharing
and reuse of software generally.

NO WARRANTY

15. BECAUSE THE LIBRARY IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE LIBRARY, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE LIBRARY "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE LIBRARY IS WITH YOU. SHOULD THE LIBRARY PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE LIBRARY AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE LIBRARY (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE LIBRARY TO OPERATE WITH ANY OTHER SOFTWARE), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Libraries

If you develop a new library, and you want it to be of the greatest possible use to the public, we recommend making it free software that everyone can redistribute and change. You can do so by permitting redistribution under these terms (or, alternatively, under the terms of the ordinary General Public License).

To apply these terms, attach the following notices to the library. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the library's name and a brief idea of what it does.>
Copyright (C) <year> <name of author>

This library is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

This library is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public
License along with this library; if not, write to the Free Software
Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA

Also add information on how to contact you by electronic and paper mail.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the library, if necessary. Here is a sample; alter the names:

Yoyodyne, Inc., hereby disclaims all copyright interest in the
library `Frob' (a library for tweaking knobs) written by James Random Hacker.

<signature of Ty Coon>, 1 April 1990
Ty Coon, President of Vice

That's all there is to it!