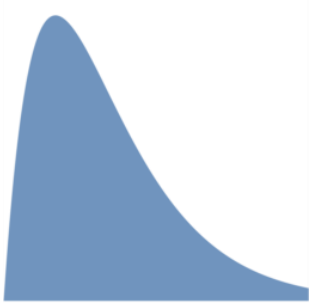


MegaStat® User's Guide

Apple MacOS Version

Copyright 2023 by J. B. Orris

Table of Contents

1. Basic Procedures ***2***

Buttons **4**

Data Selection **5**

Entering values **6**

Data Labels **6**

Output **7**

Repeat Last Option **7**

Utilities **8**

Delete output sheet 8

Start new output sheet 8

Insert descriptive information 8

ChartDataSheet utilities 9

Remove MegaStat 9

Uninstall MegaStat 9

Help/Information **10**

Help System 10

System Information 11

About MegaStat 11

2. Tutorial Examples ***13***

Example 1: Frequency distribution – selecting data **13**

Example 2: Normal Distribution – entering values; modifying output **16**

Example 3: Entering Proportions **24**

3. Reference ***25***

Descriptive Statistics **25**

Frequency Distributions **27**

Quantitative 27

Qualitative 28

Probability **30**

Counting Rules 30

Discrete Probability Distributions 31

Normal Distribution 31

t Distribution 34

F Distribution 34

Chi-square Distribution 35
Confidence Intervals / Sample Size 36
Confidence interval: mean 36
Confidence interval: p 36
Sample size: mean 37
Sample size: p 37
Sample size: mean with specified α and β 37
Hypothesis Tests.....38
Mean vs. Hypothesized Values.....38
Compare Two Independent Groups.....39
Paired Observations 39
Proportion vs. Hypothesized Value 40
Compare Two Independent Proportions 40
Chi-square Variance Test 41
Analysis of Variance 42
One-Factor ANOVA 42
Randomized Blocks ANOVA.....43
Two Factor ANOVA 44
Correlation / Regression.....45
Scatterplot 45
Correlation Matrix 46
Regression Analysis 47
Predictor values from worksheet cells.....47
Type in predictor values.....47
Select Options: 48
Select Residuals options.....49
Time Series / Forecasting 50
Trendline Curve Fit 50
Deseasonalization 51
Moving Average.....51
Exponential Smoothing.....52
Simple Exponential Smoothing 52
Two-factor Exponential Smoothing 52
Chi-Square / Crosstab.....53
Contingency Table.....53
Crosstabulation.....54
Goodness of Fit Test 55

Nonparametric Tests **56**
Sign Test 56
Runs Test for Random Sequence 56
Wilcoxon – Mann/Whitney Test 57
Wilcoxon Signed Ranks Test 57
Kruskal – Wallis Test 58
Friedman Test 58
Kendall Coefficient of Concordance 59
Spearman Coefficient of Rank Correlation 59
Fisher Exact Test 60
Quality Control Process Charts **60**
Control chart for variables (Xbar and R chart) 60
Control chart for proportion nonconforming (p chart) 60
Control chart for number of defects per sample (c chart) 60
Generate Random Numbers **61**
Appendix A. Frequently Asked Questions and Technical Information ***62***
Appendix B. Installation Instructions ***64***

MegaStat® User's Guide

MegaStat[1] is an Excel add-in that performs statistical analyses within an Excel workbook. After MegaStat is installed, it appears on the Mac Excel Data ribbon and works like any other Excel option. The purpose of this User's Guide is to introduce you to how MegaStat works. The first chapter will describe the general operating procedures and conventions that are common throughout MegaStat. The second chapter will work through a few tutorials. The Reference section shows the dialog boxes for all the options and notes briefly what data/input is expected and any unique aspects of each option.

While MegaStat is an excellent tool for learning statistics, this document focuses on using MegaStat and is not intended to teach statistics. Indeed, it assumes that you know what the various procedures do and are familiar with the terminology. It also assumes you have a basic working knowledge of Excel.

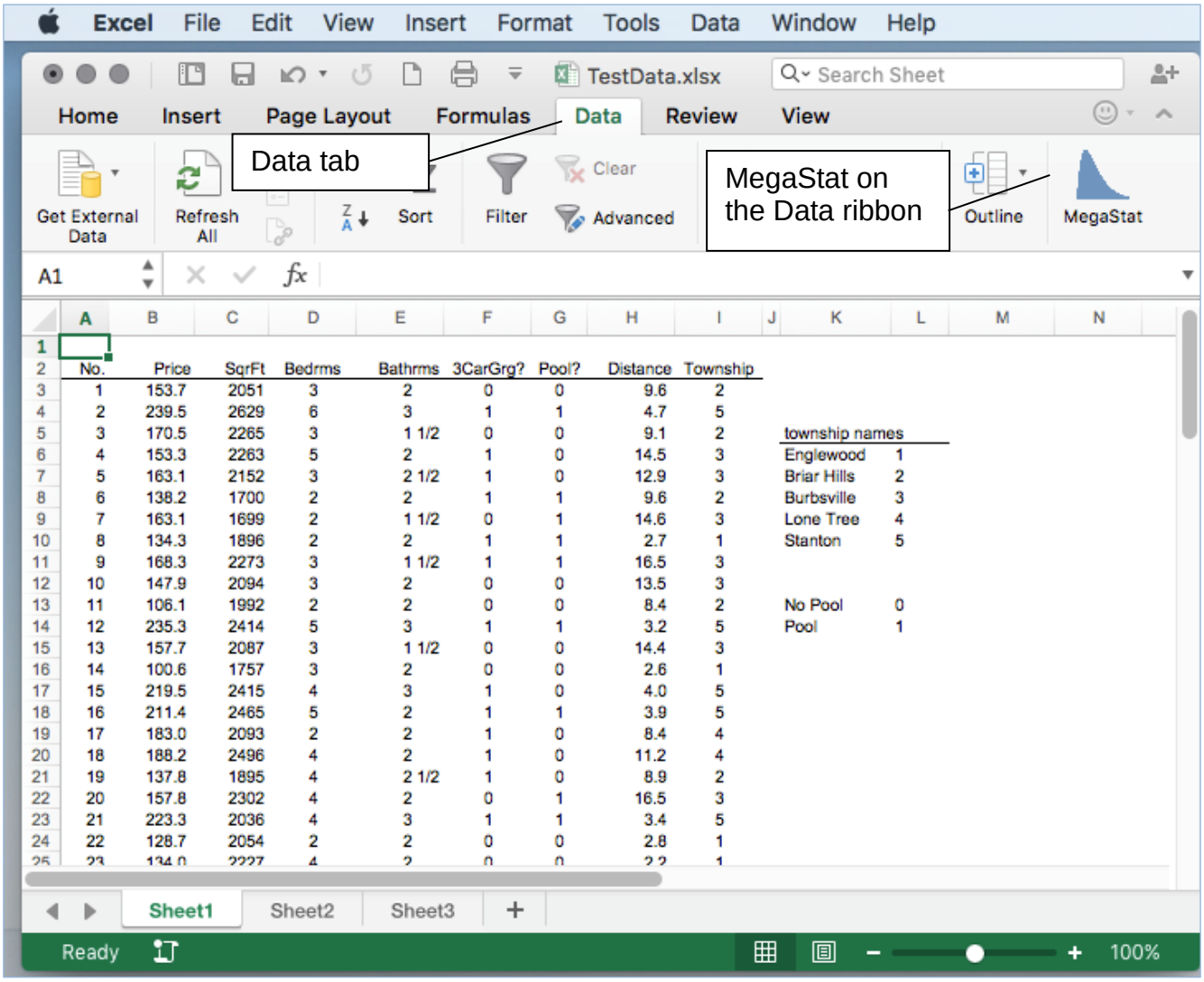
[1] MegaStat® is a registered trademark of J. B. Orris. Excel is a registered trademark of Microsoft. Mac, OS X, and macOS are registered trademarks of Apple.

1. Basic Procedures

This guide is written for MegaStat 10.4 with Mac Excel. There is also a Windows version of MegaStat that works with recent versions of Excel. The Excel versions for Mac and Windows are very similar; however, if you are using the Windows version, it would be preferable to the User's Guide for that version.

There is also a version of MegaStat for Mac Excel 2011. While this guide would be mostly relevant for that version, the menu appearance is different and some of the dialog boxes have a different appearance

After MegaStat has been installed it will appear on the Data ribbon. When you click the Data tab, your screen should look similar to Figure 1. The colors, fonts, and general appearance may be different on your computer depending on the color schemes you have selected.



No.	Price	SqrFt	Bedrms	Bathrms	3CarGrg?	Pool?	Distance	Township
1	153.7	2051	3	2	0	0	9.6	2
2	239.5	2629	6	3	1	1	4.7	5
3	170.5	2265	3	1 1/2	0	0	9.1	2
4	153.3	2263	5	2	1	0	14.5	3
5	163.1	2152	3	2 1/2	1	0	12.9	3
6	138.2	1700	2	2	1	1	9.6	2
7	163.1	1699	2	1 1/2	0	1	14.6	3
8	134.3	1896	2	2	1	1	2.7	1
9	168.3	2273	3	1 1/2	1	1	16.5	3
10	147.9	2094	3	2	0	0	13.5	3
11	106.1	1992	2	2	0	0	8.4	2
12	235.3	2414	5	3	1	1	3.2	5
13	157.7	2087	3	1 1/2	0	0	14.4	3
14	100.6	1757	3	2	0	0	2.6	1
15	219.5	2415	4	3	1	0	4.0	5
16	211.4	2465	5	2	1	1	3.9	5
17	183.0	2093	2	2	1	0	8.4	4
18	188.2	2496	4	2	1	0	11.2	4
19	137.8	1895	4	2 1/2	1	0	8.9	2
20	157.8	2302	4	2	0	1	16.5	3
21	223.3	2036	4	3	1	1	3.4	5
22	128.7	2054	2	2	0	0	2.8	1
23	134.0	2227	4	2	0	0	2.2	1

township names	
Englewood	1
Briar Hills	2
Burbsville	3
Lone Tree	4
Stanton	5

No Pool	0
Pool	1

Figure 1 Excel with the Data tab selected.

When you click on MegaStat on the Data ribbon the MegaStat menu appears (Figure 2). Most of the menu options display sub-menus (Figure 2a). If a menu item is followed by an ellipsis (…) clicking it will display the dialog box for that option.

Versions of MegaStat prior to 10.4 used dropdown menus. However, the menu structure is the same as previous versions so if you have used MegaStat before you should find this version very similar.

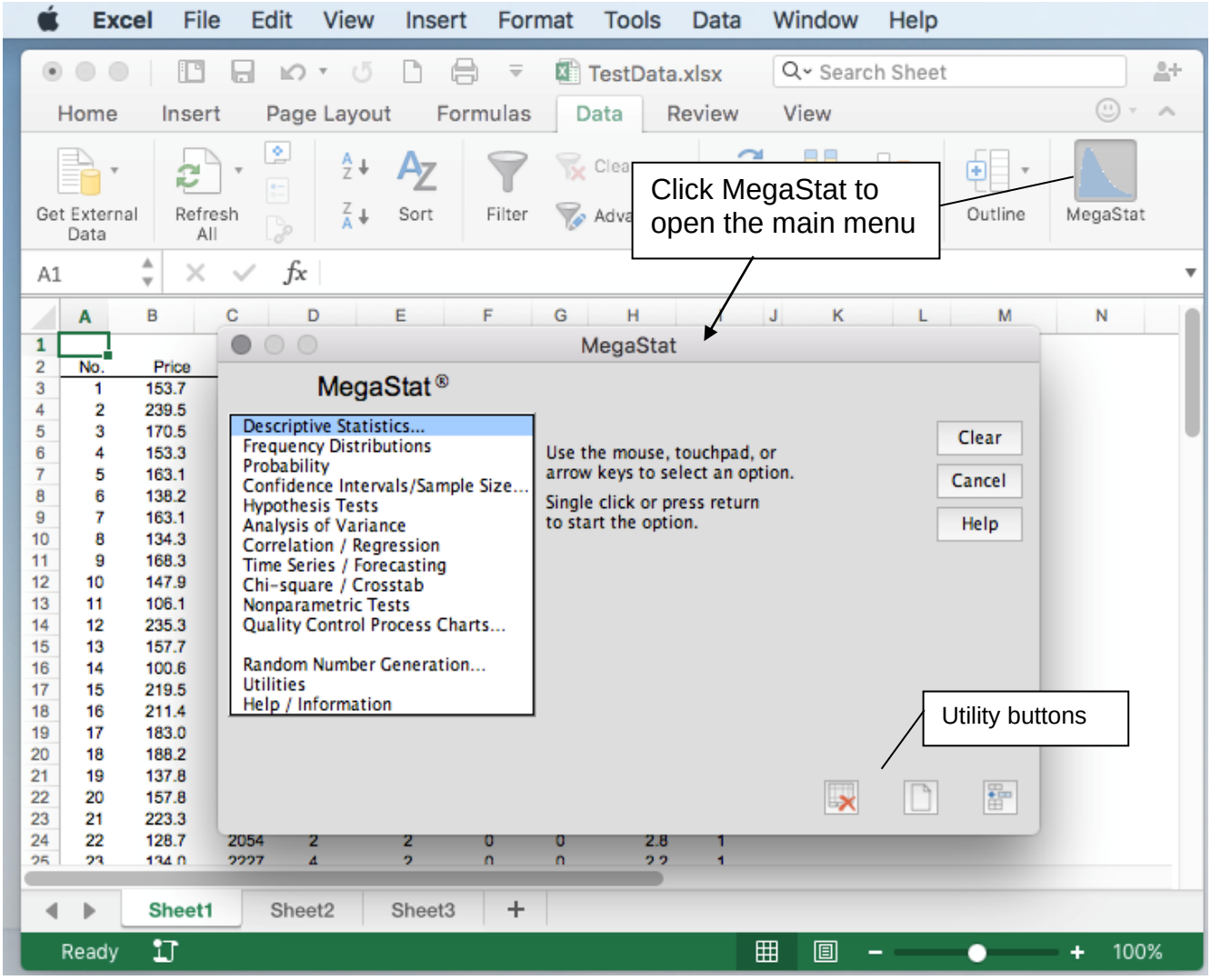


Figure 2. MegaStat main menu.

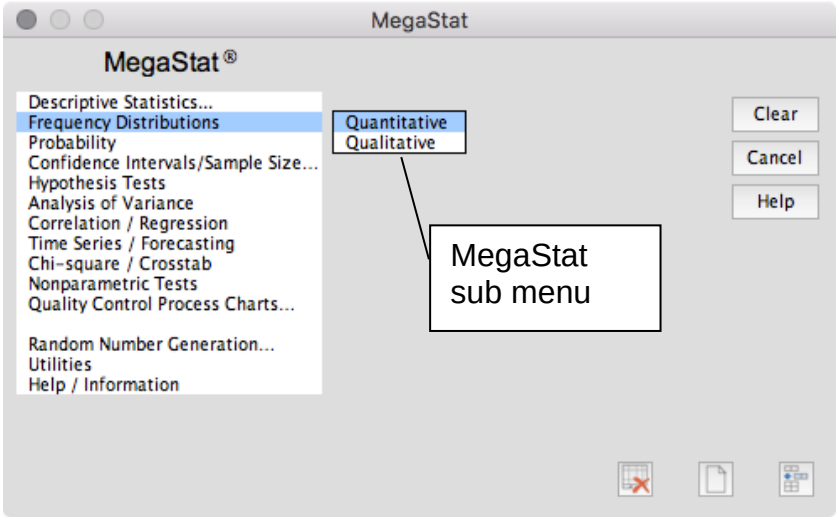


Figure 2a. MegaStat sub menu

A dialog box allows you to specify the data to be used and other inputs and options. Figure 3 shows a typical dialog box. After you have selected the data and options you click OK, the dialog box disappears and MegaStat performs the analysis.

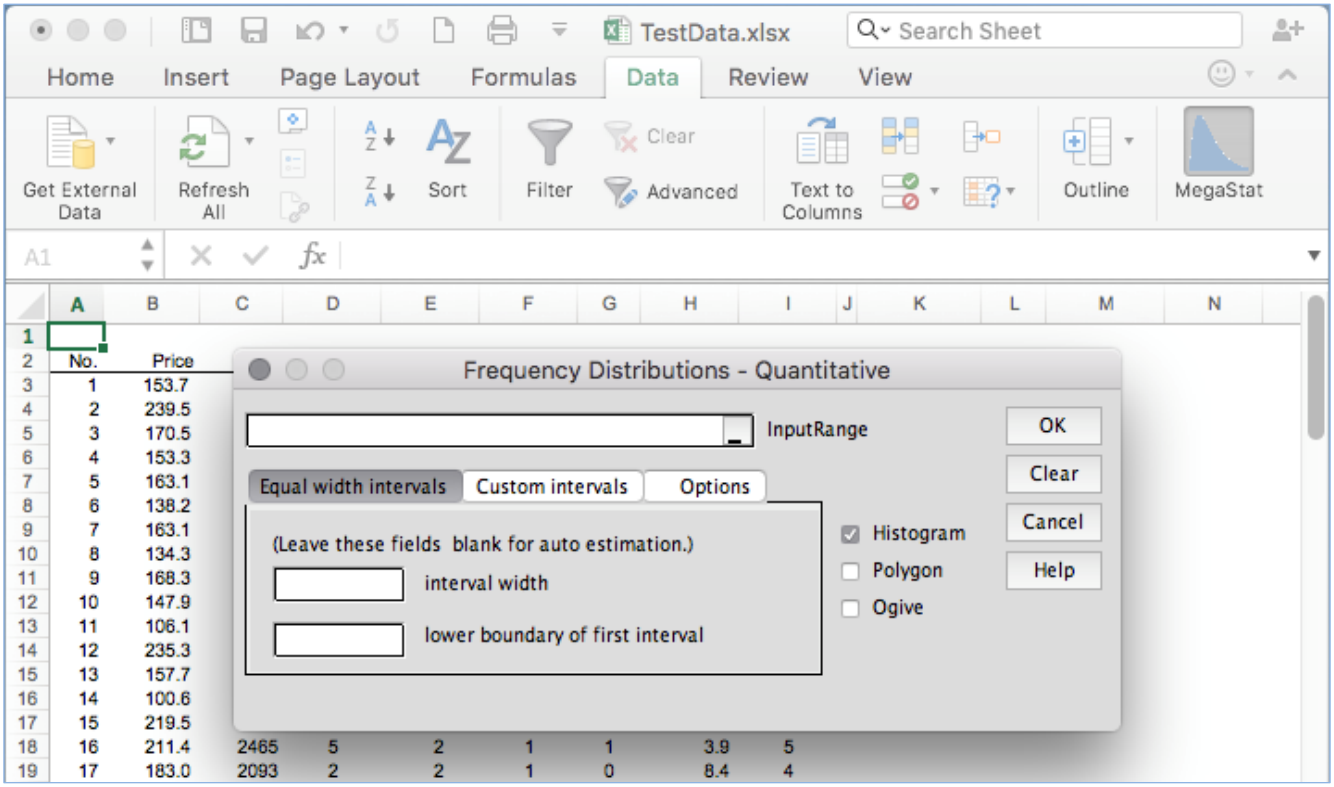


Figure 3. MegaStat dialog box

Before we look at specific dialog boxes let's take a minute to look at some items that are common to all of the options. MegaStat use is intuitive and very much like other Excel operations; however, there are some features unique to MegaStat and some ways to make using it more efficient so it will be worth your time to look at the following material

Buttons

Every dialog box has the four buttons shown on Figure 3.

OK This button could also be labeled "Calculate", "Go", "Execute" or "Do it". It tells MegaStat that you are done specifying inputs and you are turning control over to it to do its thing. First your input values are validated and then the dialog box disappears and the output worksheet is displayed with the results. When the dialog box disappears but it is still in memory and will contain the same inputs if recalled later.

Clear This button removes all input values and resets any default options on the form.

Cancel This button could be labeled "Never mind". It simply hides the dialog box. The dialog box is not cleared or removed from memory until you exit Excel. Dialog boxes do not take much memory and there is no problem with having several of them in memory. However, if you really want to unload the form, close it by clicking the close button on the title bar.

Help This button displays context sensitive help for the active dialog box. If you want to see the full Help System, use the Help selection on the main menu.

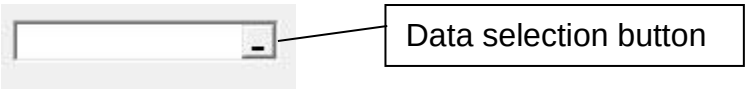
Data Selection

Most MegaStat dialog boxes have fields where you select input ranges that contain the data to be used. Input ranges can be selected four ways:

1. **Pointing and dragging with the mouse (the most common method).**

   Since the dialog box "pops-up" on the screen it may block some of your data. You can move dialog boxes around on the screen by placing the mouse pointer over the title bar (colored area at the top), clicking and holding the left mouse button while dragging the dialog box to a new location. You can even drag it partially off the screen.

   You will also notice that when you start selecting data by dragging the mouse pointer, the dialog box will collapse to a smaller size to help you see the underlying data. It will automatically return to full size when you release the mouse button. You can also collapse and uncollapse the dialog box manually by clicking the data selection button at the right end of the field. Clicking the button again will uncollapse the form. (Do not use the 'X' button to uncollapse a form.).



   If you wish to edit or delete a cell reference in an input box you need to first click the data selection button.

2. **Using MegaStat's AutoExpand feature**

   Pointing and dragging to select data can be tedious if you have a lot of data. When you drag the mouse down it is easy to 'over-shoot' the selection and then you have to drag the mouse back until you get the area correctly selected.

   AutoExpand allows rapid data selection without having to drag through the entire column of data. Here is how it works:

   - Make sure the input box has the focus. (Click in it or tab to it.) An input box has the focus when the insertion pointer is blinking in it.
   - Select one row of data by clicking in one cell of the column you want. If more than one column is being selected, drag the mouse across the columns.
   - The data range will expand to include all of the rows in the region where you selected one row when you do one of the following:
     - Double-click over the input field
     - Right-click over the input field
     - Left-click the label next to input box.

   With a little practice you will find this is a very efficient way to select data. The only time you cannot use it is when you want to use a partial column of data. You should also be aware that the AutoExpand stops when it finds a blank or non-numeric cell; thus any summations or other calculations at the bottom of a column would be selected. It is good practice to leave a blank cell at the bottom of each column before inserting formulas or text.

Note: When using the above methods of data selection, you may select variables in an alternate sequence by holding the **command** ⌘ key while making multiple selections and then do the AutoExpand.

3. **Typing the name of a named range**

   If you have previously identified a range of cells using Excel's name box, you may use that name to specify a data range in a MegaStat dialog box. This method can be very useful if you are using the same data for several different statistical procedures. Range names are not case sensitive.

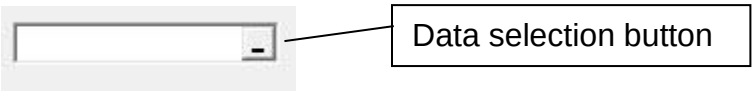
4. **Typing a range address**

   You may type in any valid Excel range address, e.g. B5:B43. This is the least efficient way to specify data ranges but it works.

Entering values

If an input box requires a **single value**, you may do one of the following (make sure the insertion cursor is blinking in the box):

- Type a value into the box.
- If an input box has a data selection button (as shown below) that means that, in addition to typing in a value, you may also select an existing value by clicking a cell.



  Click on any Excel cell that contains a value. When you click on a cell, the cell address is shown in the input box. If you double-click the input box, the address will change to the value in the cell.

- Type any formula that you could be entered into a cell. You do not have to type the '=' sign as you would in an Excel cell.
- Type a cell address, e.g. B6, or the name of a named cell.

Data Labels

For most procedures the first cell in each input range can be a label. **If the first cell in a range is text it is considered a label**; **if the first cell is a numeric value it is considered data**. If you want to use numbers as variable labels you must enter the numbers as text by preceding them with a single quote mark *e.g.* '2. Even though Excel stores times and dates as numbers, MegaStat will recognize them as labels if they are formatted as time/date values.

If data labels are not part of the input range, the program automatically uses the cell immediately above the data range as a label if it contains a text value.

If an option can consider the entire first row (or column) of an input range as labels, <u>any</u> numeric value in the row will cause the entire row to be treated as data.

If the program detects sequential integers (1,2,3…) in a location where you might want labels it will display a warning message otherwise the rule is: **text cells are labels**, **numeric cells are data**[2].

[2] An exception is the Crosstabulation option that can count text data.

Output

When you click OK on a MegaStat dialog box it performs some statistical analysis and needs a place to put its output. It looks for a worksheet named Output. If it finds one, it goes to the end of it and appends its output; if it doesn't find an Output worksheet it creates one. MegaStat will never make any changes to the user's worksheets, it only sends output to its Output sheet.

MegaStat makes a good attempt at formatting the output but it is important to remember that the Output sheet is just a standard Excel worksheet and can be modified in any way by the user. You can adjust column widths and change any formatting that you think needs improvement. You can insert, delete and modify cells. You can copy all or part of the output to another worksheet or to another application such as a word processor.

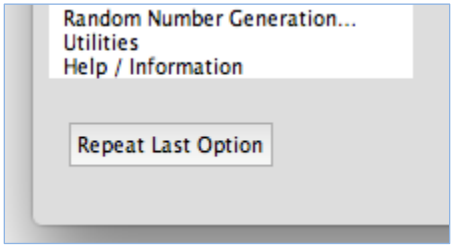
MegaStat charts get their values from cells on the Output sheet (or one of your worksheets in the case of the Scatterplot). You can click a chart and select 'Source Data' to see what values are being displayed.

When the program generates output it adjusts column widths for the current output. If you have previous output from a different option already in the Output sheet, the column widths for the previous output may get messed up. You can attempt to fix this by manually adjusting the columns widths or by always starting a new output sheet.

The Utilities menu has options for deleting the Output sheet or making a copy of it and starting a new one.

Repeat Last Option

Once you have performed a MegaStat option, this menu selection will allow you to re-display the last dialog box without having to go through the menu selections.



This can be handy if you need to make a change or when you need to repeat the same operation with the different data sets.

Utilities

The Utilities menu contains some items that perform useful functions.

Delete output sheet

This option deletes the current Output sheet. It will present a warning message because there is no way to recover a sheet once it is deleted. You may delete the Output sheet without the warning message by right-clicking the button on the main menu dialog box (see below).

Start new output sheet

If there is an existing Output sheet it will be renamed Output(2) so that your next output will be on a fresh Output sheet. You can rename Output(2) to whatever you wish by double-clicking the name tab.

You can also start a new Output sheet by using Excel to rename the existing Output sheet (right-click the Output worksheet tab).

Insert descriptive information

This option is used for identifying output. It will insert rows with labels for Description, Name, Workbook, Date/Time and Environment (version of the operating system, Excel, and MegaStat).

You then use the adjacent cells to type in the Description and Name information. Figure 9 in the next chapter shows an Output sheet after clicking this option.

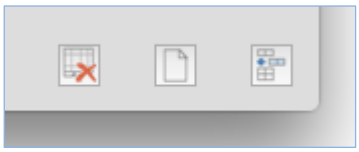
The Workbook line will show the name and location of the active workbook and the name of the worksheet where the information is being inserted. The Time entry with show the current time/date and the version of the operating system, Excel, and MegaStat. Any of the inserted information may be edited.

This option is not limited to MegaStat output sheets – it can be used on any Excel worksheet. The cells will be entered at the active cell.

If you inadvertently insert the information where you didn't intend to, select the five rows and delete them or Cut/Copy them to another location.

The first tutorial example in the next section (Figure 9) shows an example of inserting descriptive information.

The Delete, Start new output sheet, and Insert information utilities can also be accessed from buttons on MegaStat main menu:



If you right-click the buttons the operation is invoked without displaying a message box. Just remember that there is no undo operation for a deleted output sheet.

ChartDataSheet utilities

In order to display a graphical output (*e.g.,* a chart) Excel must reference values in a worksheet. If the values are not available as a part of the output sheet, MegaStat stores them in a hidden worksheet called "ChartDataSheet_". If MegaStat needs a ChartDataSheet it creates one unless one already exists, in which case it appends its values to the end.

The following options use ChartDataSheets:

Descriptive Statistics – Boxplot

Regression Analysis – Plot residuals by X values

Regression Analysis – Normal Probability Plot

Quality Control Process Charts

In general, you do not need to be concerned with ChartDataSheets and probably will not even be aware that they exist; however, the following utilities exist for advanced users.

Show

ChartDataSheets exist only to provide values to charts and thus the output is not labeled. However, at the top of each output section is a label telling what type of chart it is used for and a time/date stamp. Under the corresponding chart there is also a time/date stamp so you can associate the data with a chart.

If you change or delete any of the values on ChartDataSheet the corresponding chart will be changed.

Hide

This will hide the ChartDataSheet after viewing it.

Delete

You would use this option if you wanted to delete a ChartDataSheet that no longer has any associated charts. If you delete a ChartDataSheet that has an existing chart, the chart will still exist but will not have any values plotted. There is no Undo so make sure before you click OK.

You can also right-click the worksheet tabs to Unhide, Hide, and Delete the ChartDataSheets.

Remove MegaStat

The option displays a message box telling you how to remove MegaStat from the Data ribbon by doing: Tools > Excel Add-ins… > and unchecking MegaStat

To restore the 'MegaStat' to the Data ribbon by: Tools > Excel Add-ins… > and then checking the MegaStat option that you will see in the list of available add-ins.

Uninstall MegaStat

This menu item does not actually uninstall MegaStat. It displays a dialog box prompting you on how to do the uninstallation process. You should remove MegaStat from the Data ribbon before uninstalling

Uninstalling is the process of using Finder to remove the MegaStat folder from the Applications folder – either delete the folder or move it to another location.

Help/Information

Help System

This option displays the full MegaStat help program as shown in Figure 4. Help works by opening the default web browser (usually Safari). If the browser is already open a new tab will be created. If you see just part of the Help window, click it to bring it to the front. If you have the Excel window maximized or if the Excel window is larger than the browser window nothing may appear to happen when you click Help – you will need to click Safari (or your default browser) in the Dock to view help and click Excel to get back to MegaStat after viewing the help.

The MegaStat dialog box may initially be in the foreground but it will go back of the help screen if you click the help window. When you are done viewing help you can close the browser or just the MegaStat Help tab. To get back to MegaStat without closing the browser, click the Excel window or click Excel in the Dock.

The 'How MegaStat works -- general operating procedures" section contains all the information in this tutorial. You can click specific topics or search for an item by clicking on Index. Click Using Help for details on using the help system. The Help button on dialog and message boxes will display context sensitive help.

Figure 4a. MegaStat Help System

System Information

This form displays technical information regarding software versions and file locations. If you click the Insert button, the information will be placed in the current Output sheet as shown in Figure 4c.

About MegaStat

This option displays current version information. There is also a link for the MegaStat website. This links will only work if you have an active Internet connection and your system is setup to properly respond to Internet links. Figures 4b shows About and System information displayed in the same window.

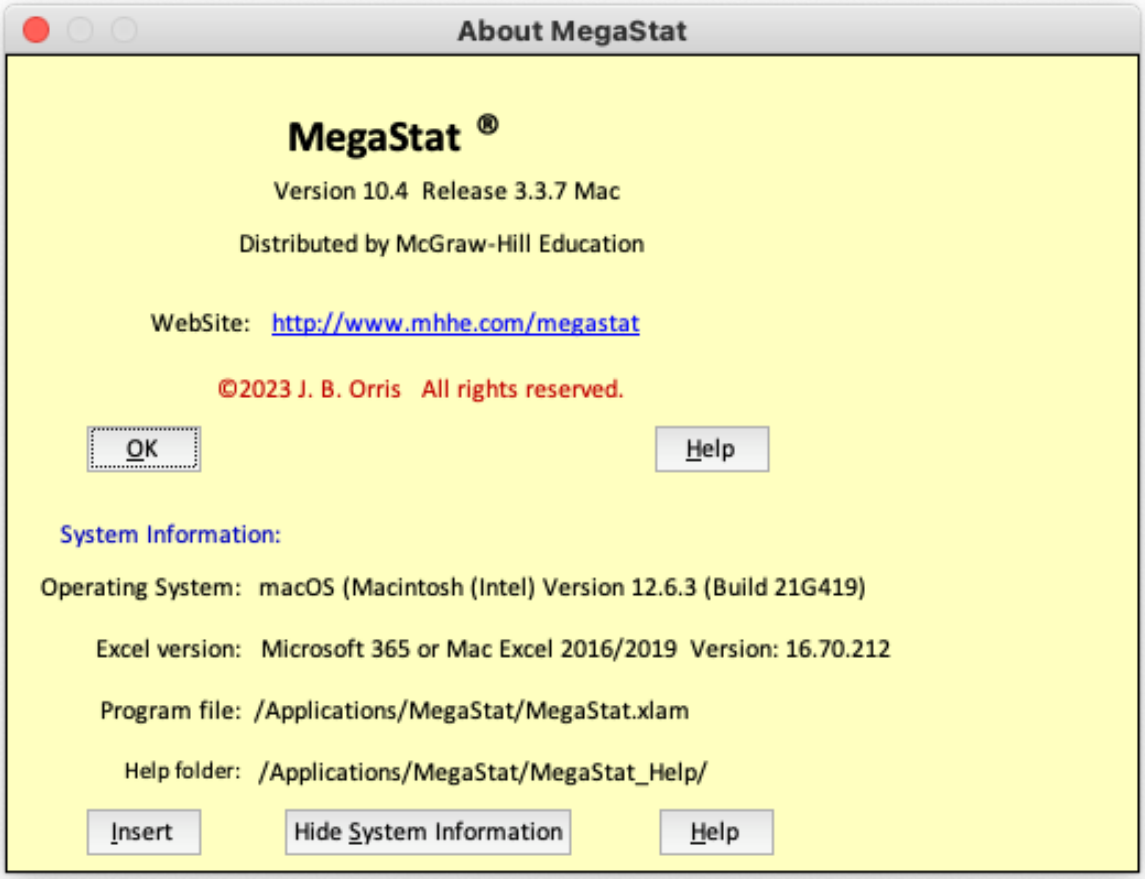


Figure 4b. About MegaStat and System Information.

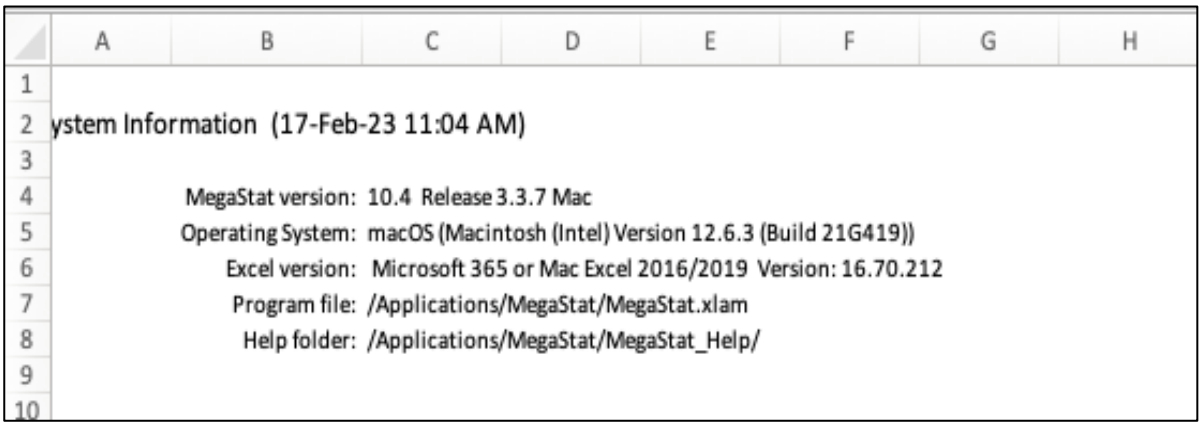


Figure 4c. System Information in Output sheet by clicking 'Insert'.

2. Tutorial Examples

Although MegaStat performs many different statistical options the various dialog boxes all work the same way and have standard Excel objects (input boxes, buttons, checkboxes, etc.). This chapter will work through a few detailed examples and will point out a few things that are unique to MegaStat. The next chapter will provide a reference source for the various options.

Example 1: Frequency distribution – selecting data

The first tutorial example will perform a quantitative frequency distribution on the Price variable of the Sheet1 worksheet of the Testdata.xlsx workbook. If you want to work through this example, start Excel and open Testdata.xlsx and click on the Sheet1 worksheet tab.

The steps in this font are what you would do to work through the tutorials.

a. Open the Quantitative Frequency Distributions dialog box by clicking: MegaStat → Frequency Distributions → Quantitative.

b. Click the Help button on the MegaStat main menu dialog box. Review the Help screen and then exit Help.

Since we want to do the frequency distribution on the Price variable we need to select all the data in column B. This example will illustrate the use of the AutoExpand feature although you could get the same result with any of the other methods of data selection.

c. Make sure the insertion cursor is blinking in the input box. If necessary, click in the box with the left mouse button.

d. Click anywhere in column B, e.g. cell B6

e. Place the mouse pointer over the input box and double click the left mouse button activate AutoExpand.

You can also activate AutoExpand by right-clicking the input box or left-clicking the 'Input range' label. Figure 5 shows the dialog box at this point.

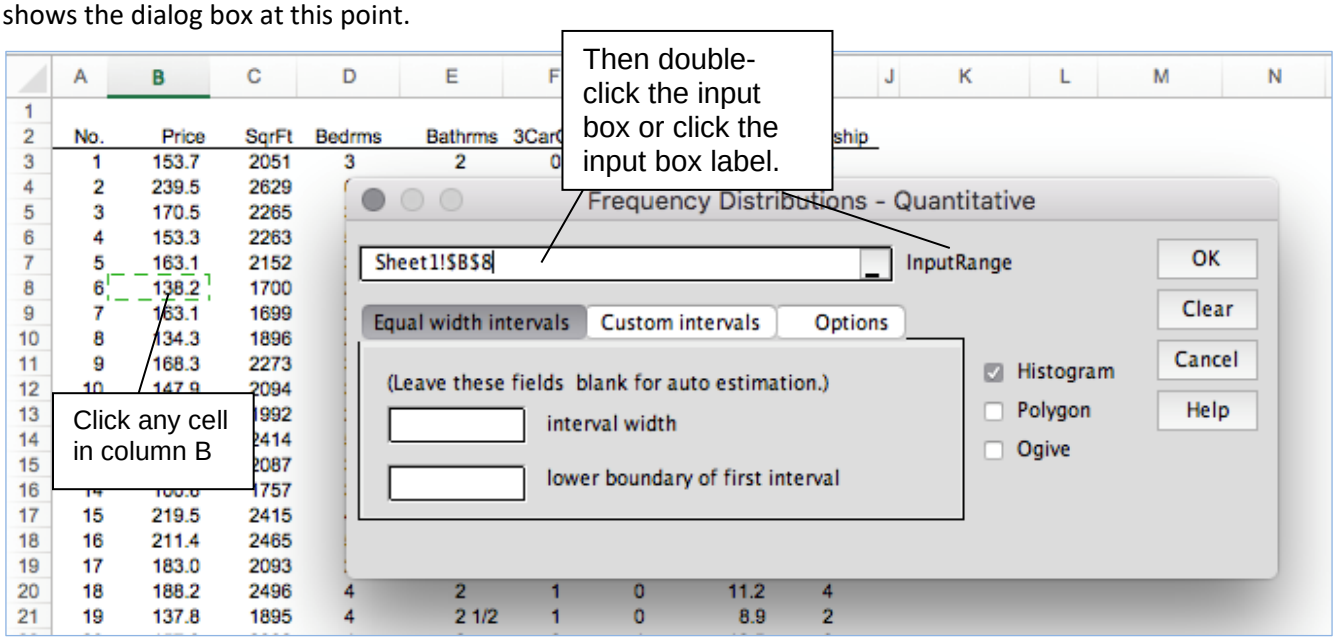


Figure 5. Illustration of AutoExpand

After clicking the label, the dialog box appears as shown in Figure 6 with all of the data in column B being selected. The selection expands up/down until it encounters a blank cell or the top/bottom of the column.

When using AutoExpand make sure you do not have sums or other non-data values in the last cell of the range.

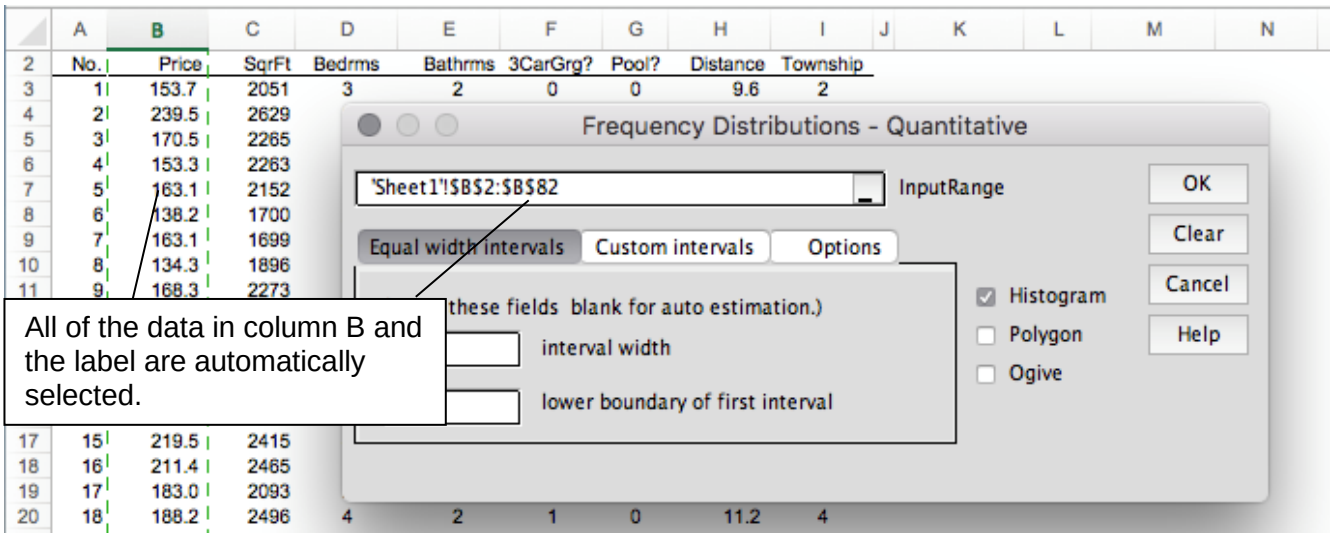


Figure 6. All of the data in column B and the label are selected with AutoExpand.

f. Type 50 in the interval width box. Click OK.

Determining the proper number to enter for the interval width is something you will learn in your statistics course. If you leave either input box empty, the MegaStat will attempt to calculate an appropriate value. Prior to clicking OK, the dialog box will look like Figure 7.

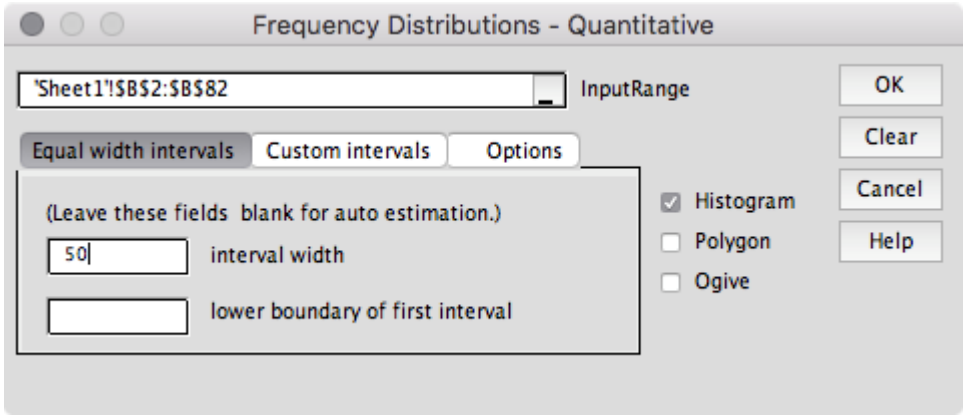


Figure 7. Completed dialog box waiting for OK click.

After you click OK the dialog box disappears and MegaStat does the requested calculations. A new Output worksheet (shown in Figure 8) is created and displayed. If there were already an Output sheet MegaStat would have appended the new output to the end.

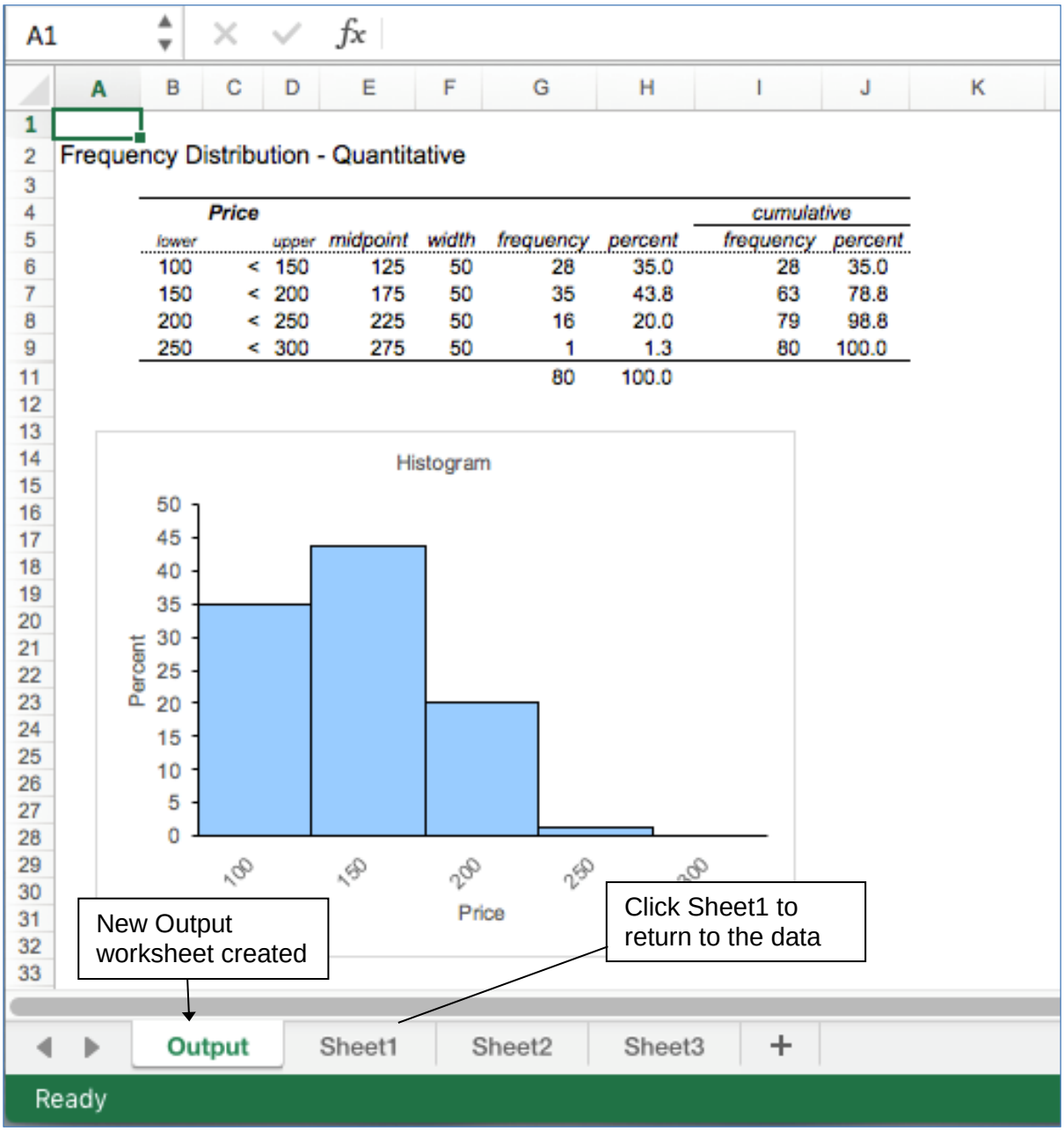


Figure 8. MegaStat Output sheet

The amount displayed on your screen will be different depending on the size of your screen. You may need to use the vertical and/or horizontal scrollbars to view the entire output.

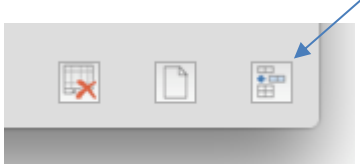
The Output sheet is just an Excel worksheet and it can be formatted and manipulated just like any other worksheet. You may change the font style / size / color; increase / decrease decimal places; add annotations, etc. However, you do need to be careful if you delete part of the output. For example, if you delete the midpoint cells, the Frequency Polygon cannot be plotted.

Also note that the charts are Excel charts and may be edited also to change colors, axis scaling and anything else. One thing you should routinely do is change the generic title (e.g. 'Histogram') to something more meaningful and relevant to the data.

Insert descriptive information

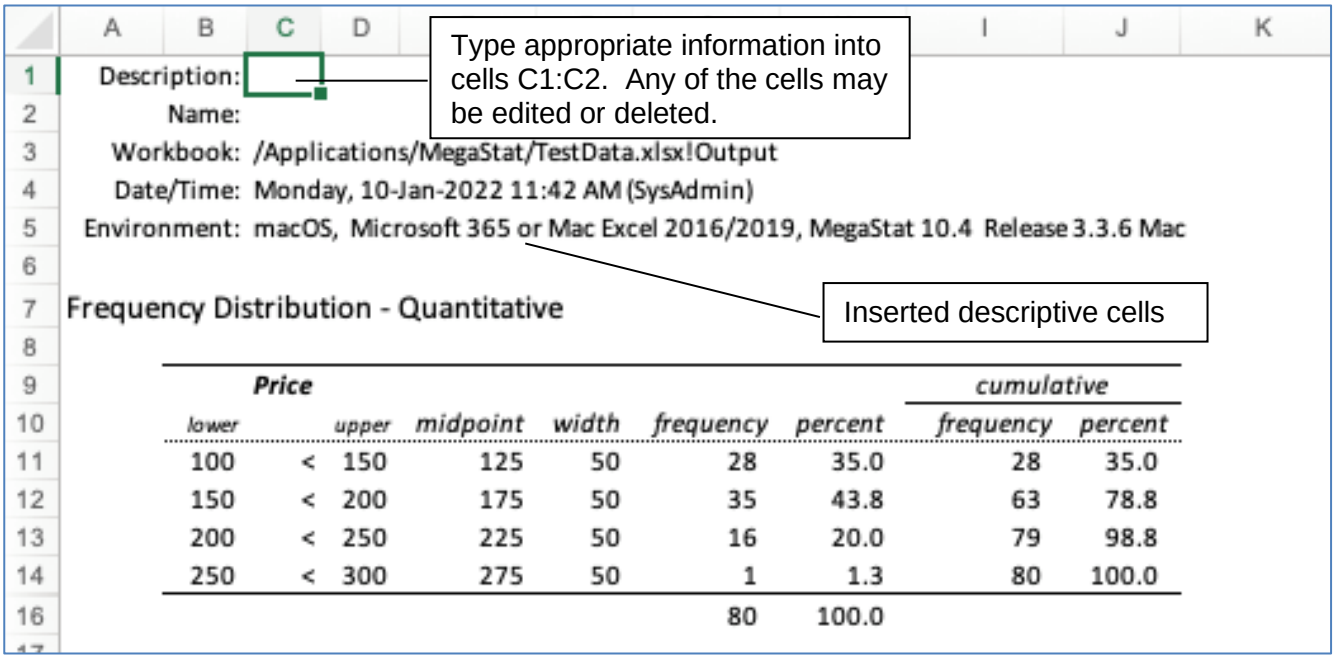
If you are submitting MegaStat output as part of a report or homework project, you can have MegaStat insert descriptive cells.

g. Click in cell A1 and then click MegaStat → Utilities → Insert descriptive information. You could also use the utility button on the main menu.
(End of tutorial steps for Example 1)



This will insert rows with descriptive labels (as shown in Figure 9) and you can type in the appropriate information and edit and/or delete the Data Source and Date/Time.

x



Description:	
Name:	
Workbook:	/Applications/MegaStat/TestData.xlsx!Output
Date/Time:	Monday, 10-Jan-2022 11:42 AM (SysAdmin)
Environment:	macOS, Microsoft 365 or Mac Excel 2016/2019, MegaStat 10.4 Release 3.3.6 Mac

Frequency Distribution - Quantitative

Price							cumulative	
lower		upper	midpoint	width	frequency	percent	frequency	percent
100	<	150	125	50	28	35.0	28	35.0
150	<	200	175	50	35	43.8	63	78.8
200	<	250	225	50	16	20.0	79	98.8
250	<	300	275	50	1	1.3	80	100.0
					80	100.0		

Figure 9. Inserted descriptive information.

Example 2: Normal Distribution – entering values; modifying output

Most MegaStat options work with data; however, some options use other types of input. For example, when using probability distributions, you need to specify the appropriate input values. The tutorial example will use the normal distribution dialog box but the concepts apply also to other MegaStat options.

To run this option, you must specify values for z, mean, and standard deviation. The default values for the mean and standard deviation are 0 and 1 which means the input would be a z value; however, you may specify any mean and standard deviation.

For example, using the TestData.xlsx file, what is the probability that a house would be less than 2500 square feet (assuming a normal distribution)? Let's first use the Descriptive Statistics option to find the mean and standard deviation.

a. Open TestData.xlsx

b. Run MegaStat → Descriptive Statistics with the default options for the SqrFt variable.

c. Locate the cells that contain the mean and standard deviation.

d. (optional) Run MegaStat → Frequency Distributions → Quantitative on the SqrFt variable to verify that the distribution is approximately normal.

e. Select MegaStat → Utilities → Start new output sheet. This will rename the Output sheet Output(2) so our next output will be on a fresh sheet.

f. With the Output(2) worksheet selected, open the Normal Distribution dialog box by clicking: MegaStat → Probability → Normal Distribution.

The dialog box appears as shown in Figure 10a.

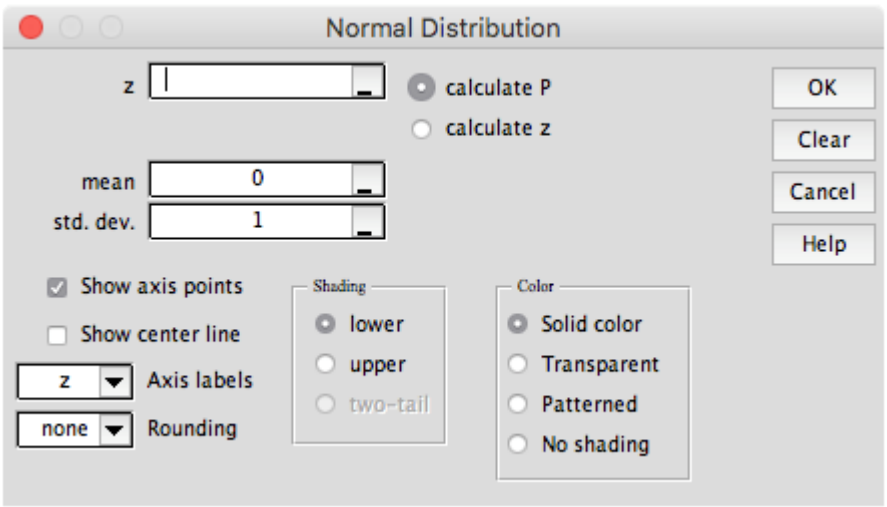


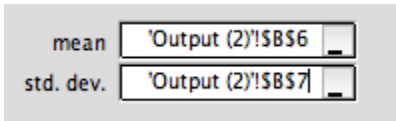
Figure 10a. Normal distribution dialog box.

You could just type the mean and standard deviation in the boxes but let's get the values from the cells.

g. Click in the mean box and then click the cell that contains the mean.

h. Click in the standard deviation box and then click the cell that contains the standard deviation.

The boxes should now look similar to this depending which cells had the mean and standard deviation:



It is OK to leave the boxes like this; however, if you double-click each of the boxes the actual numbers will be shown and the dialog box will appear as shown in Figure 10b.

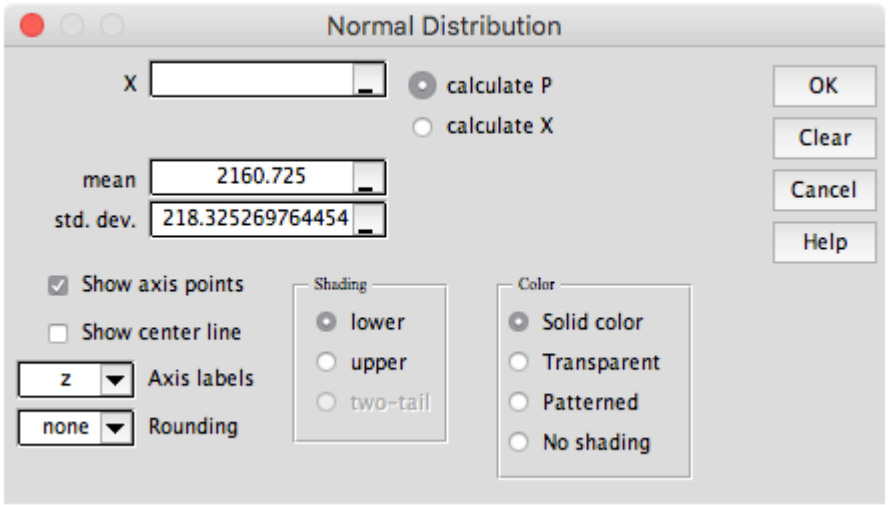


Figure 10b. Normal distribution dialog box with mean and standard deviation.

Notice that an advantage of selecting the cell rather than typing in the values is we get the full accuracy of the values since you would probably have typed in the rounded values. However, if you did not want all of the decimal places you could click the cells and edit the values.

Note that the input labels that were 'z' in Figure 10a are now labeled 'X' in Figure 10b since the mean and standard deviation are no longer 0 and 1.

If a cell contained the X value, you could also click in that cell but we will just type it in. We will also specify the X axis labels and round the calculated z value to two places so the output will correspond to table values.

i. Click in the input box for X and type 2500.

j. Click the dropdown arrow for Axis labels and select 'X'

k. Click the dropdown arrow for Rounding and select '2'.

l. Click OK and the output will appear as shown in Figure10c.

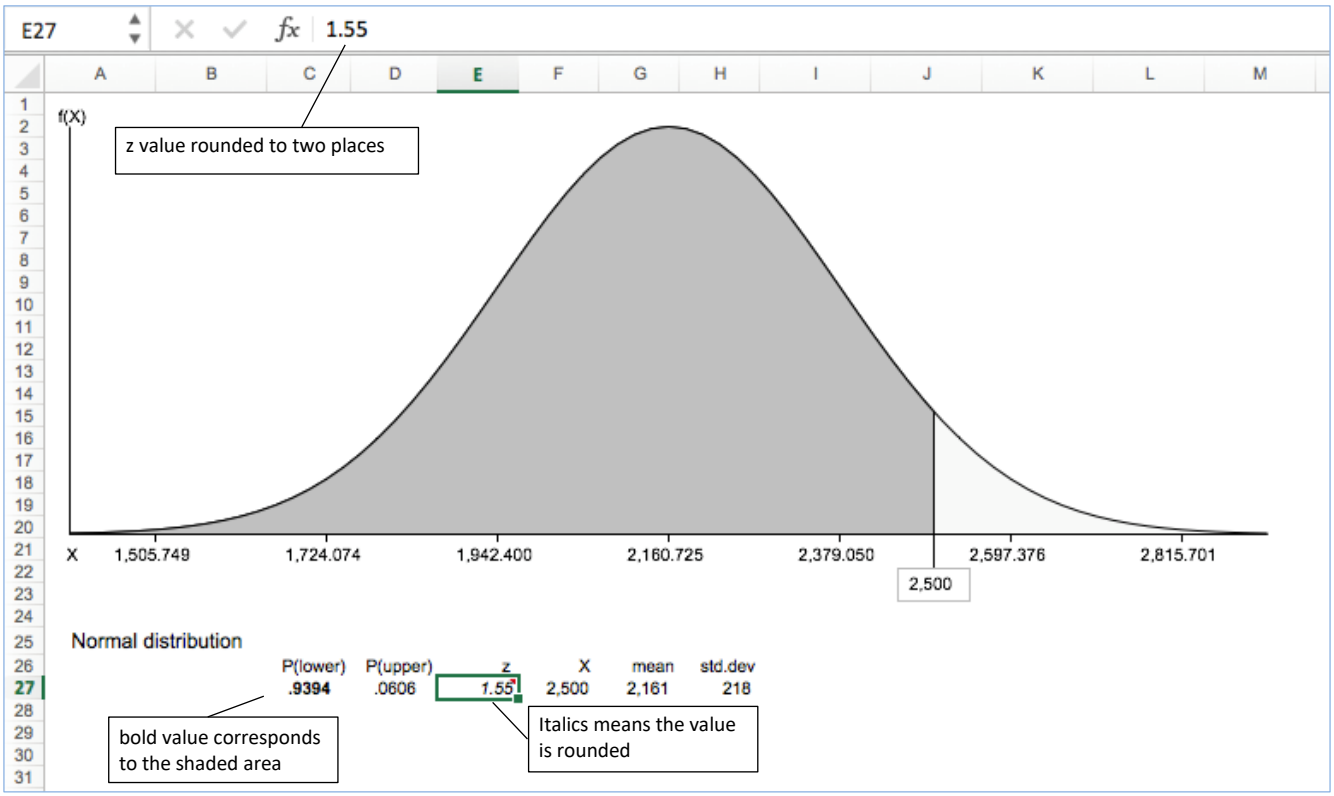


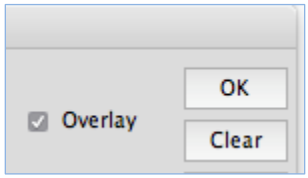
Figure 10c. Output from the normal distribution option.

Note that in addition to the graphical output the input values and the calculated values are also shown. The probability shown in bold corresponds to the shaded area. The z value in italics means that it is a rounded value, not just formatted, *i.e.*, if you click in the cell you will see the value is actually 1.55000000000. If you don't want the bold and italics formatting, you can click in those cell and change it, indeed you can change any of the output. For example, you may want to click in the cells with the mean and standard deviation and increase the number of decimal places to correspond to axis values.

We will now calculate the value for a 2000 square foot house and superimpose the output on this display.

m. Click MegaStat → Repeat Last Option (or MegaStat → Probability → Normal Distribution).

Note there is new checkbox near the OK button:



When the Overlay option appears it is checked by default which means the next output will be superimposed on the previous output. If you uncheck it, a new output will be created. For this example, we will leave it checked.

n. Enter 2000 into the 'X' box.

o. Select 'Transparent' in the Color box and click OK.

The new output is shown in Figure 10d.

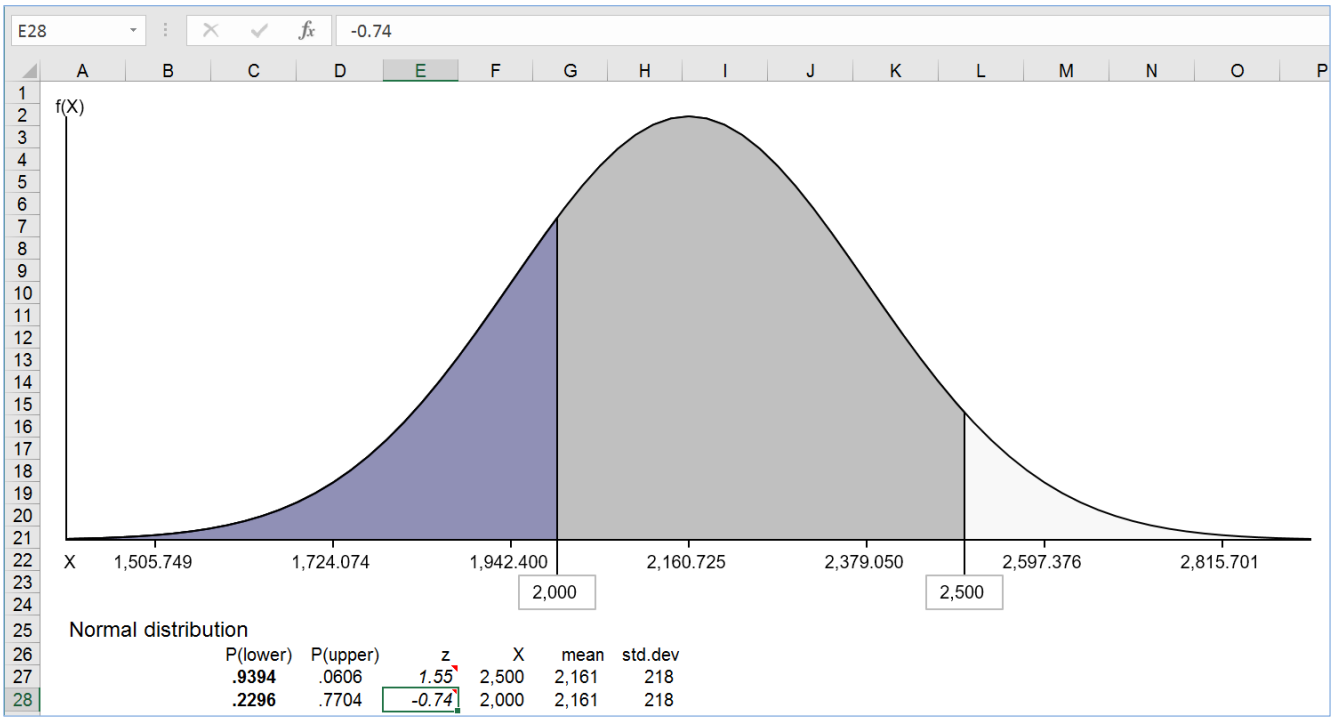


Figure 10d. Normal distribution with two areas shown.

Although the new area is transparent it would be more evident if the original area was patterned. We could go back and re-do the output and specify patterned instead of solid color; however, we will show how you change it on the output (and in the process see how you can do all kinds of fancy and colorful output).

Each of the shaded areas (and the axis labels) are separate graphical objects that can be edited. For this example, we will change the larger shade area to brick pattern.

p. Click the larger shaded area (the area between 2000 and 2500).

q. Click the Shape Format tab. Click the dropdown menu on Shape Fill.

The screen will look like Figure 10e.

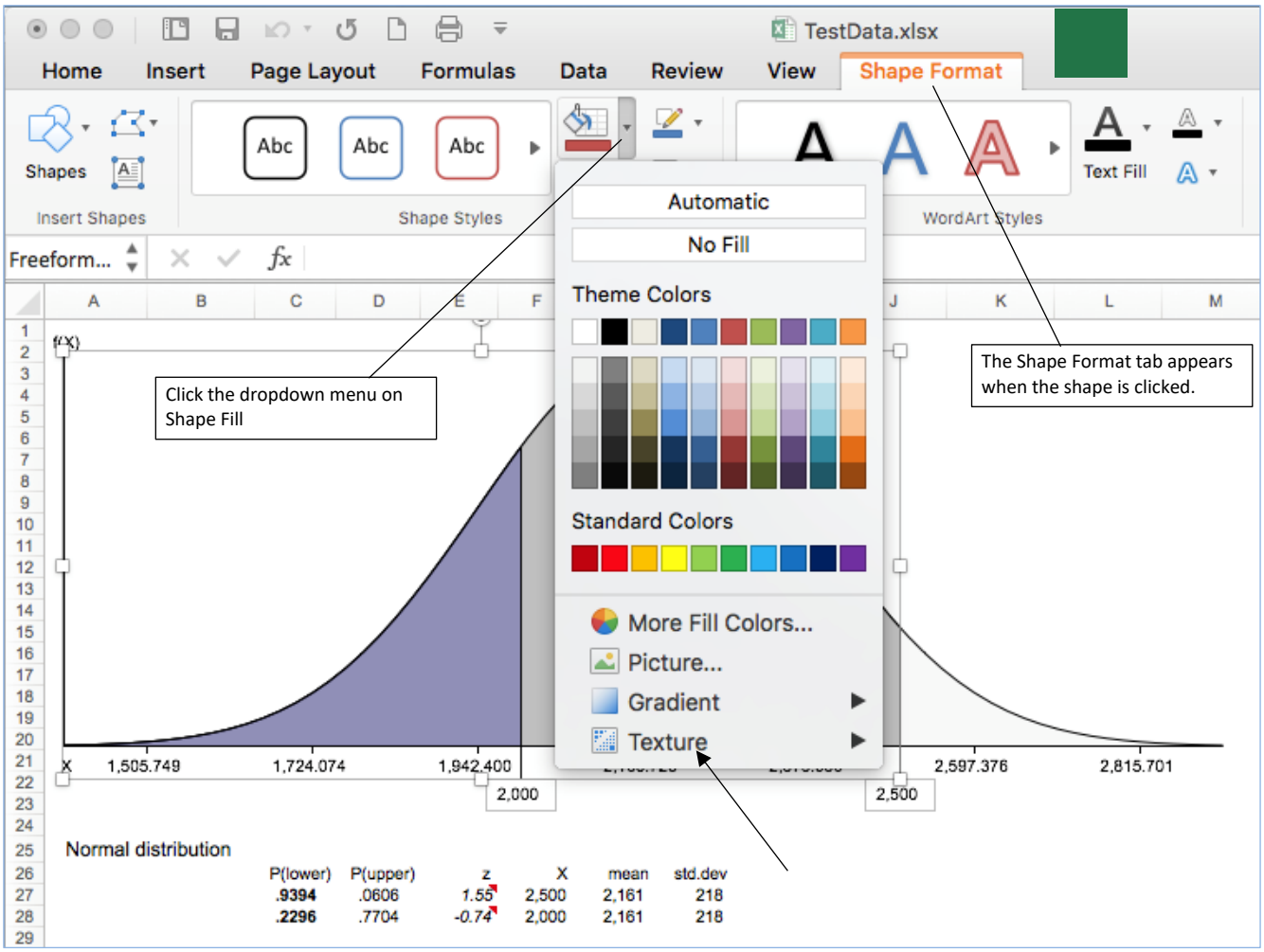


Figure 10e. Shape Fill menu.

r. Click Texture → More Textures and the Format Shape task pane will appear. Click Pattern fill and select a pattern.

The Format Shape task pane will appear as shown in Figure 10f.

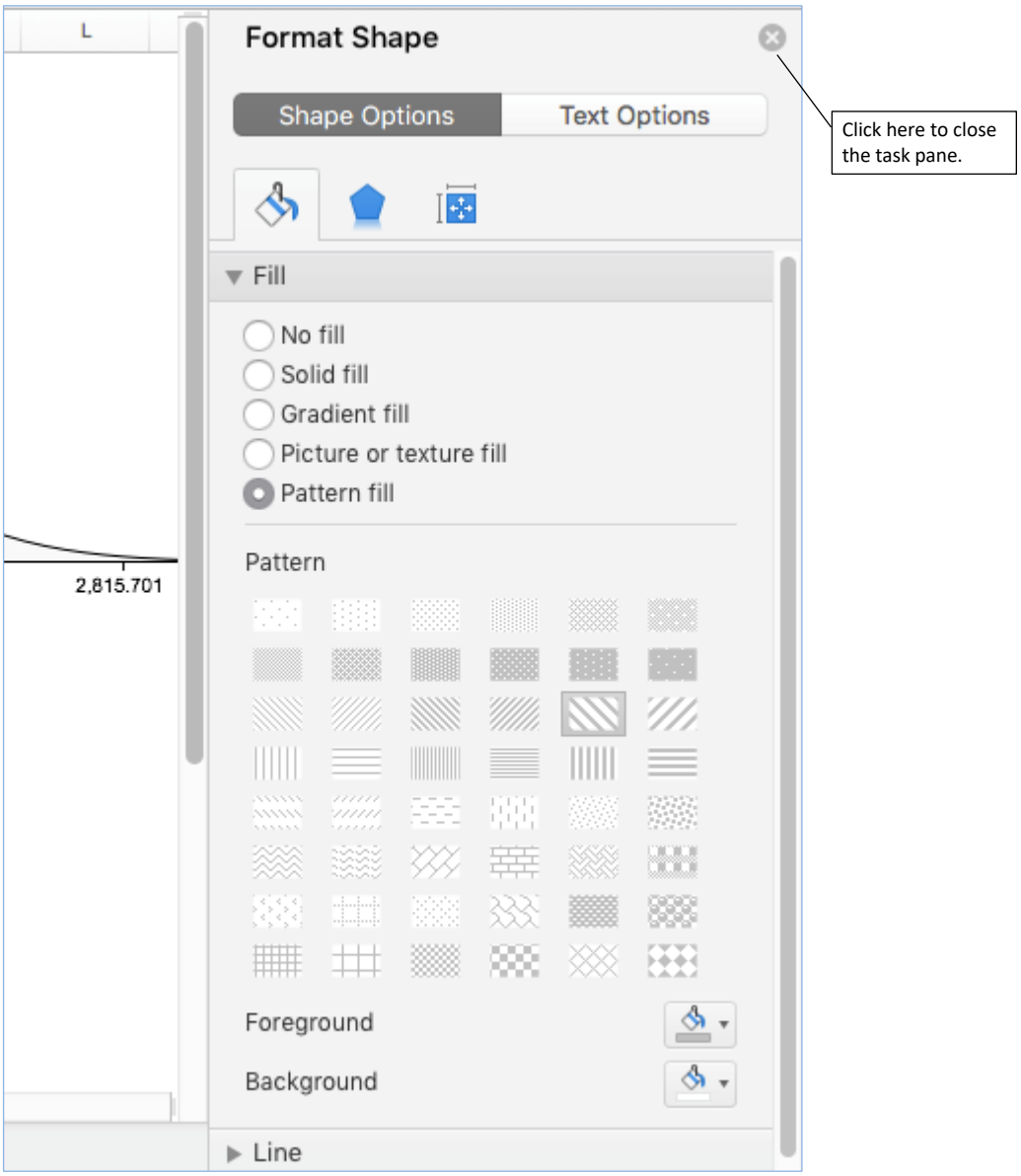


Figure 10f. Format Shape task pane.

Note that the Format Shape dialog box can also be used to change many other properties of the shape. You can even put a digital picture in the areas. Try some of them.

s. Close the task pane.

t. Calculate the area under the curve between 2000 and 2500 by clicking in cell C29 and entering the formula: =C27-C28.

(End of tutorial steps for Example 2.)

Your output will now look like Figure 10g.

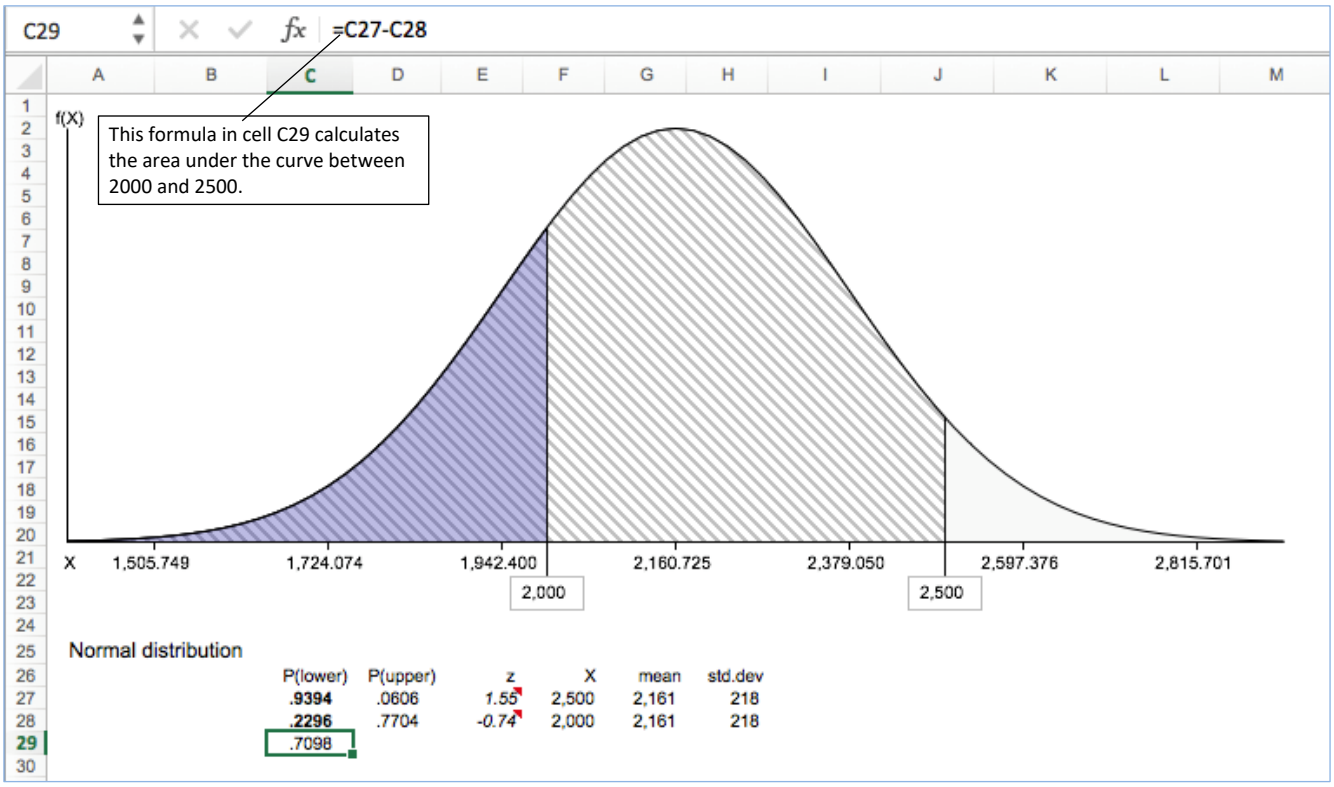


Figure 10g. Output modified to show a patterned fill.

Notice how the patterned fill with the transparent overlay emphasizes that the probability for 2500 goes all the way to left.

Example 3: Entering Proportions

Several MegaStat dialog boxes require that you enter proportions. This particular example calculates the confidence interval if the sample proportion is 11/38. You could enter the decimal value of 0.28947368421; however, this would be time consuming and error prone. It would be better to just enter the fraction as shown in Figure 11.

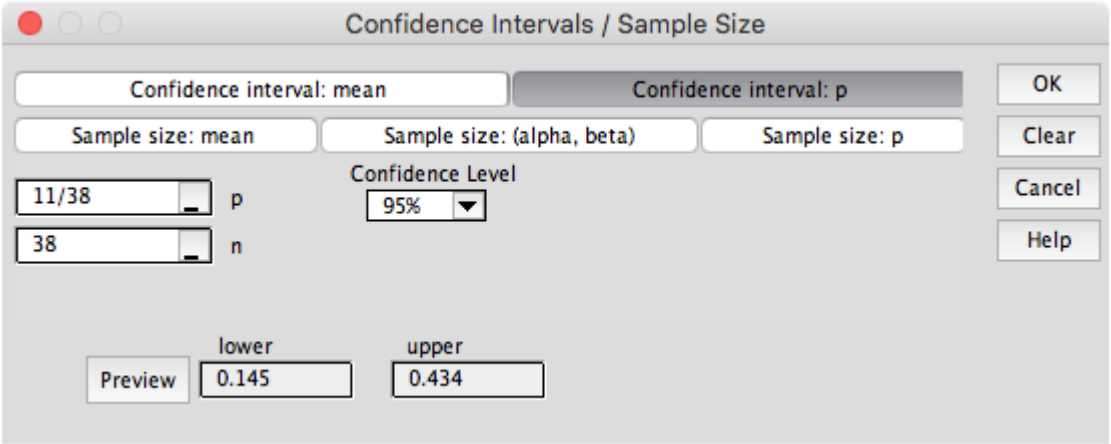


Figure 11 Proportion entered as a fraction.

An even better method is to just enter the numerator portion of the proportion and the program will divide the n value entered in the n box. When you enter a value in the p box that is greater than or equal to 1, the label automatically changes to x and the program knows that is supposed to divide by n (see Figure 12).

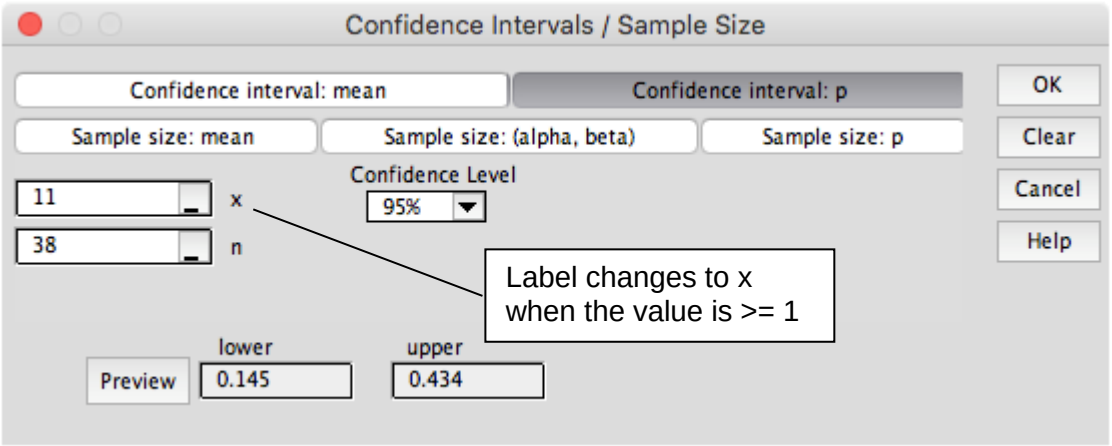


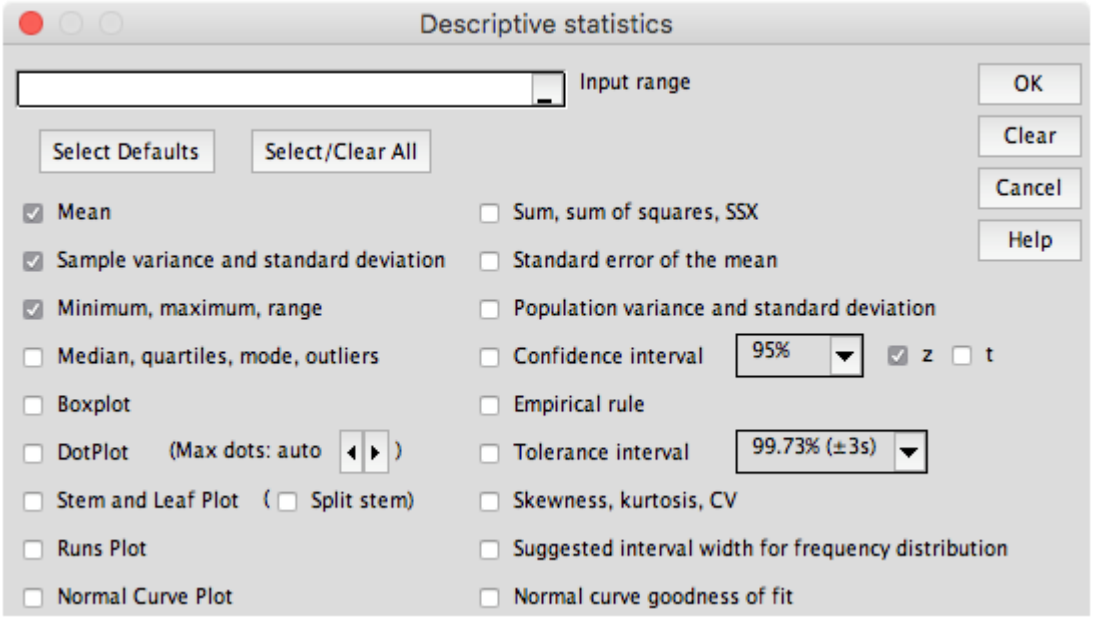
Figure 12 Proportion entered by typing only numerator.

This method works whenever MegaStat requires a proportion. If a worksheet cell contains a proportion, you also can click on that cell to select the proportion.

3. Reference

This section lists MegaStat options and briefly discusses any issues relevant to the option. Each dialog box is displayed; however, only features and procedures that are not self-apparent are discussed.

Descriptive Statistics



The Select Defaults button selects the options shown. Select/Clear All toggles 'Select All' and 'Clear All' options.

Note 1: The option that calculates **outliers** defines outliers and extremes as follows:

Let: Q1 = 25th percentile
Q3 = 75th percentile
H = Q3 – Q1 (the interquartile range)

An <u>outlier</u> is defined as any value less than Q1 - 1.5*H or greater than Q3 + 1.5*H

An <u>extreme</u> is defined as any value less than Q1 – 3.0*H or greater than Q3 + 3.0*H

Note 2: Boxplot fences

If the data contains outliers or extremes (see Note 1) they are plotted individually on the boxplots and the fences are displayed as dashed vertical lines. Fences are the values that define outliers and extremes.

Note 3: Boxplot charts

The data values needed to display the chart(s) are stored in ChartDataSheets as discussed in the Utilities section. If you paste the chart into another application, you might want to do a Paste Special → Picture that does not require a link to these values.

Note 4: Confidence intervals

Confidence intervals may be calculated using z-values, t-values, or both by checking the appropriate box(es). The default is the z-value. You may select the confidence level from the dropdown menu or you can type a value.

Note 5: Normal curve goodness of fit

This option uses the mean and standard deviation of the data to determine the intervals boundaries that would divide a normal distribution into equal-probability intervals. The number of intervals is determined by Sturge's rule: $1 + \log_2(n)$. The row header for the observed values shows the z-value corresponding the upper end of each interval.

The program then calculates the frequency distribution. The expected frequencies are n/(number of intervals). If the data are normally distributed each interval should have the same number of observations.

Interpretation: A larger p-value indicates that the observed distribution closely matches a normal distribution. As the p-value gets smaller the fit is not as good. If the p-value is under .05 you would reject the hypothesis that the data could have come from a normal distribution.

Note: The sample size for the first variable is used to determine the number of intervals for all of the variables. If calculating the test for multiple variables and some of the other variables contain substantial amounts of missing data you should run the test on them individually.

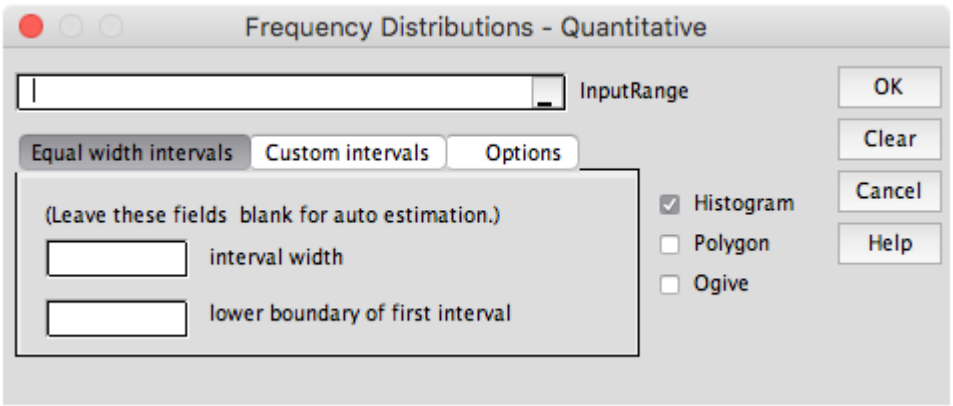
Frequency Distributions

Quantitative

The basic operation of this option is described in the Tutorial Example section above.

If the interval width or lower boundary boxes are left empty, MegaStat will attempt to calculate appropriate values: To estimate the interval width it first calculates the range of the data excluding any outliers or extremes. It then determines the interval width by the method discussed in David P. Doane "Aesthetic Frequency Classifications", The American Statistician, November 1976, Vol 30, No. 4.

If your data includes outliers the resulting frequency distribution will probably have some empty intervals in order to reach the outliers. You might exclude some data or consider custom intervals. The dialog box for frequency distributions with equal width intervals is shown and discussed in the Tutorial Example section above. However, when you have a wide range of data, perhaps with some outliers you would click the Custom Intervals tab and the dialog box would appear as follows:



Use this option to specify unequal width or open-ended intervals. Select the worksheet cells (called the 'bin range') containing the interval boundaries. Each cell will be the lower boundary of an interval except for the last cell, which will be the upper boundary of the last interval. The values must be selected in a single column.

Example:

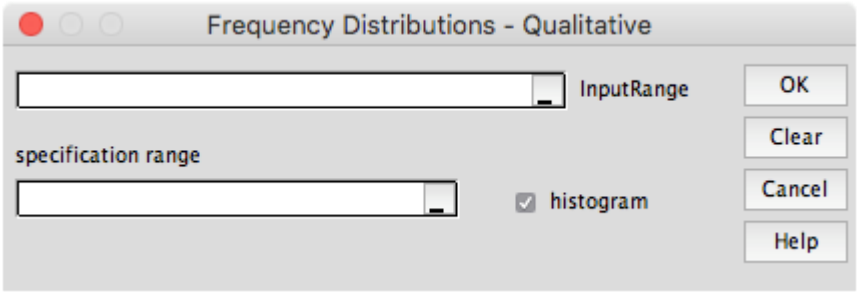
This bin range

0
100
200
1000
4000
10000

would create the these intervals:

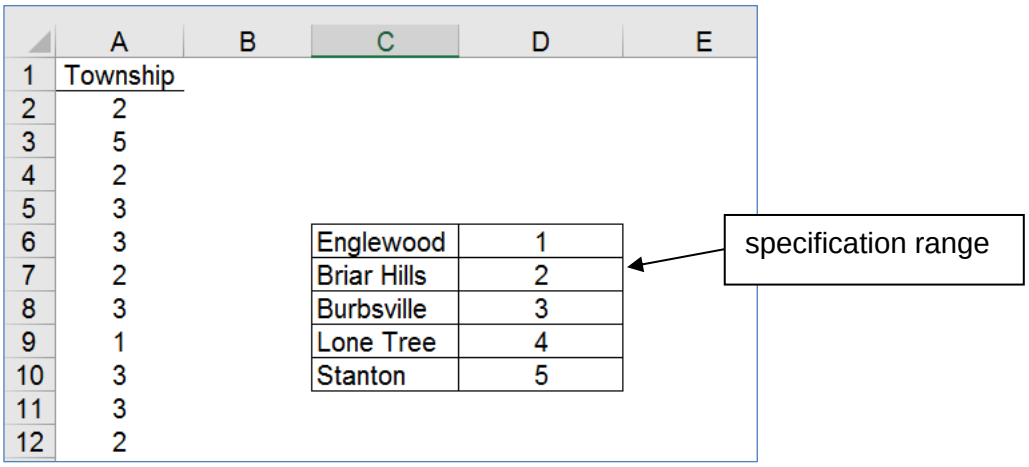
lower	upper
0	99.99
100	199.99
200	999.99
1000	3999.99
4000	10000.00

Qualitative



A qualitative frequency distribution is used to count the number of occurrences of specified data values. A **specification range** is used to specify what values are to be counted.

Suppose you had a variable representing five townships coded 1-5. If you wanted to include the name of the townships on the output you would select cells C6:D10 as the specification range as shown below. (Note: the cell borders shown are not required as a part of the specification range.)



Suppose townships 1, 3 and 4 were in the eastern part of the city and townships 2 and 5 were in the western part. If you wanted to count the number of homes in the East and West region you would set the specification range as shown below. Values listed on the same row are counted in the same category. If no label is specified, the left hand data value would be used as the label.

Any data values not listed in the specification range are ignored.

	A	B	C	D	E	F
1	Township					
2	2					
3	5					
4	2					
5	3					
6	3		East	1	3	4
7	2		West	2	5	
8	3					
9	1					
10	3					

specification range

MegaStat also allows qualitative data in the data range. For example, if the Garage variable was coded Yes and No the specification range would be as follows:

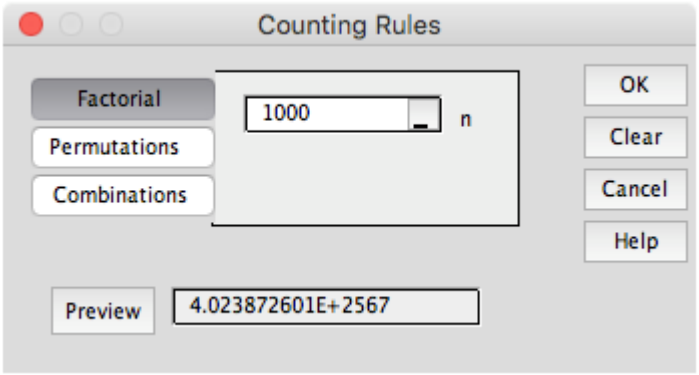
	A	B	C	D
1	3CarGrg?			
2	No			
3	Yes			
4	No		Yes	
5	Yes		No	
6	Yes			
7	Yes			
8	No			
9	Yes			

specification range

If you enter qualitative data as text if must be typed exactly the same way every time and it is case sensitive: Yes and yes would be two different responses. You could get around this by typing Yes and yes on the same line in the specification range but generally it would be better to enter numeric values such as 0 and 1. These could also be used as indicator variables in regression.

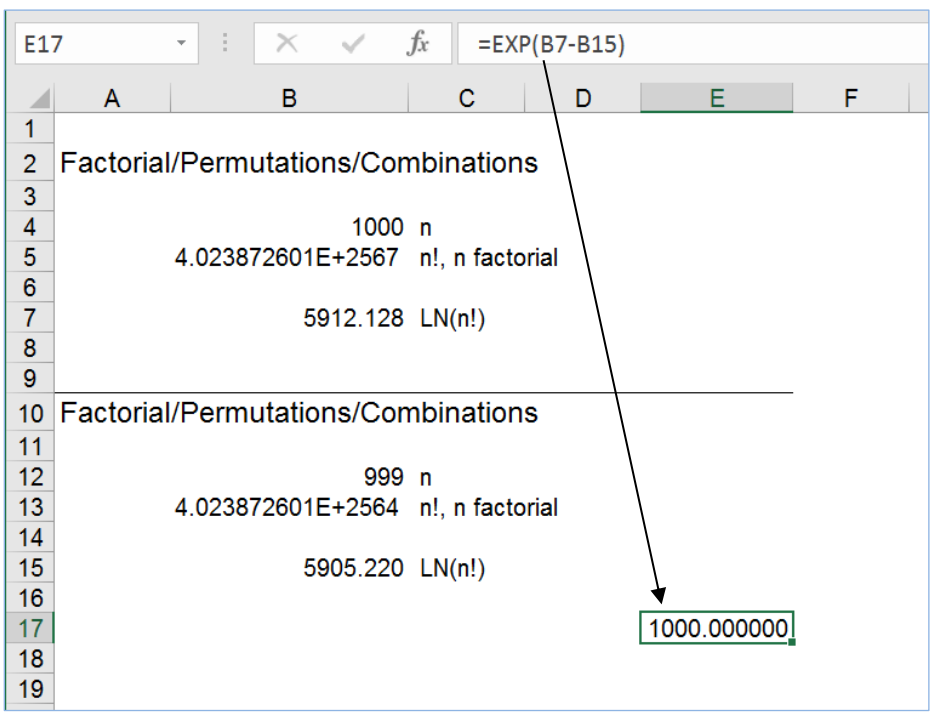
Probability

Counting Rules

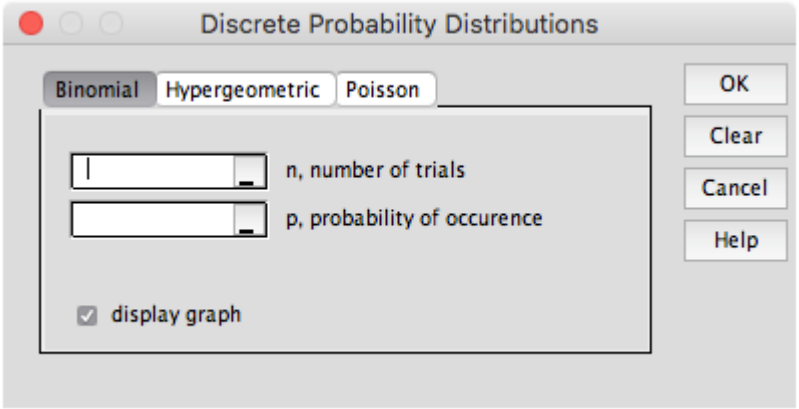


The largest number Excel can handle is approximately 1.0E306, i.e., a number with 306 places to the right of the decimal. That is certainly a large number but factorial/permutation/combination calculations often generate larger values. The largest factorial that can be displayed as a number is 170!. Values larger than that are calculated but the results are shown in text format. The Preview example above shows 1000!.

Technical note: The output also includes the natural log of the answer. You may use this value for further computations and testing. For example, to verify that the factorials are correct, recall that n!/(n-1)! = n, therefore exp(ln(n!)-ln((n-1)!) should equal n, e.g. exp(ln(1000!)-ln(999!) = 1000. Values larger than 10,000,000! start to show a little error but the percent error is very small. These computations are shown below in cell E17 below:



Discrete Probability Distributions

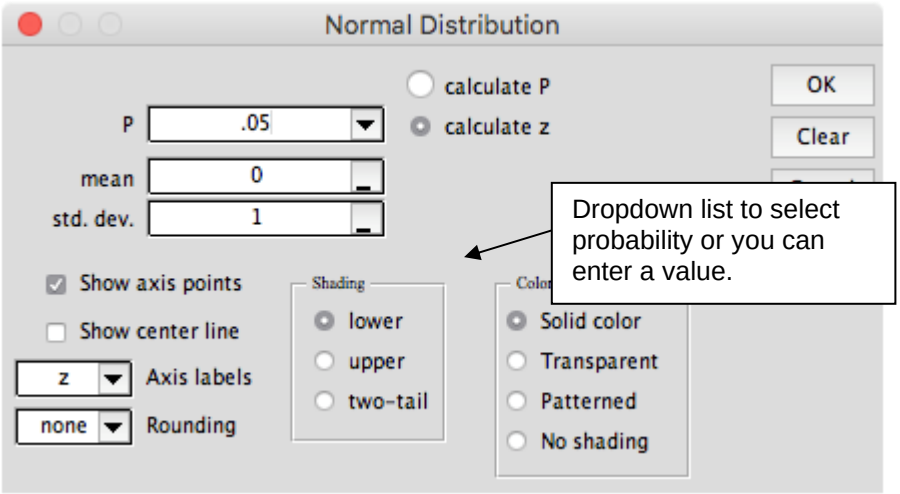


Enter the values required by each distribution. The Help file gives details regarding size limitations.

The output is formatted to five decimal places so a value that shows as .00000 is not truly zero, it is just zero out to at least five decimal places. Like any MegaStat output, the formats can be changed. You will need to use scientific notation to display very small probabilities.

Normal Distribution

The normal distribution program gives a graphical display of the normal distribution. Example 2 in the Tutorial chapter shows how to calculate normal curve probabilities. The dialog box below shows how it would be set up to calculate the z-values associated with a .05 two tail probability selected from the dropdown box (.025 in each tail of the distribution) with patterned output.



You may type in a probability value if the value you want is not in the list.

After you do an initial output, an Overlay checkbox will appear near the OK button. If you leave it checked, when you click OK the output will be superimposed on the previous output. You may superimpose several outputs. If you are planning outputs that will be covering each other, it will work best to make the first one patterned and subsequent ones transparent.

Notes regarding the options:

Show axis points – This shows input values associated with the shaded areas. The units will be z or X depending on the Axis labels option below.

Show center line – The center line is also a graphics object that can be selected and moved or edited; for example, you may want to make it heavier or change its color.

Axis labels – This option determines the labeling for the tick marks and the axis points. The options are z, X, and none.

Rounding – The options are none, 2, 3, or 4; or you may type a value from 0 to 16. If you select this option, probabilities will be calculated with the z-value rounded to the selected value. Use rounding of 2 to match most normal curve tables.

Shading – The two-tail option is only enabled when the input box is P. The input p-value is divided by two and each half is placed in one tail of the distribution.

Color – The solid color is a medium gray; the transparent color is a grayish blue that will show patterns underneath. Two or more overlaid transparent areas will show darker blue; however, it would be better to get a darker blue by editing the area as shown in Example 2 in the tutorial section.

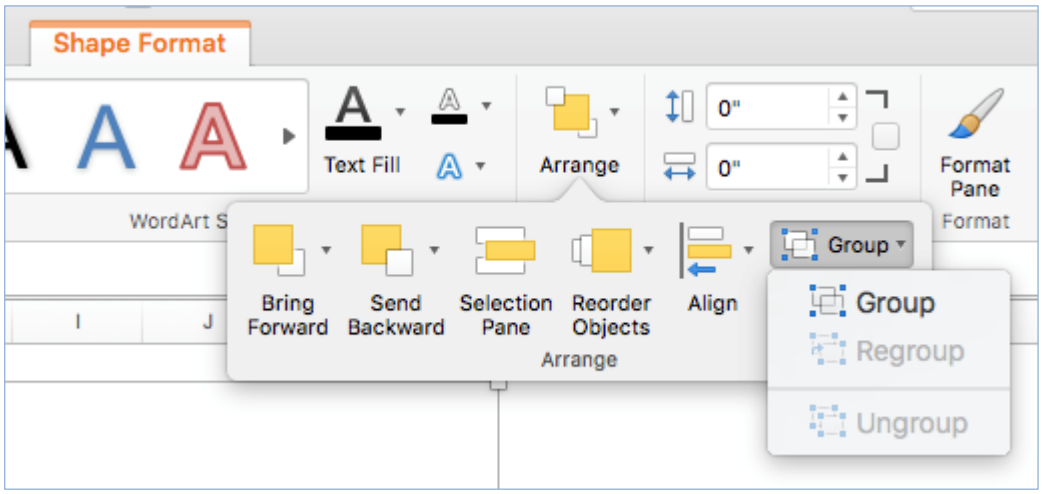
Other notes:

Post output editing: MegaStat creates the normal distribution output by displaying graphical objects (shapes and textboxes). If you click any of the areas on the output a Drawing Tools tab will appear at the top of the screen. As shown in Example 2 in the tutorial section, with a little practice you can create very interesting graphical displays.

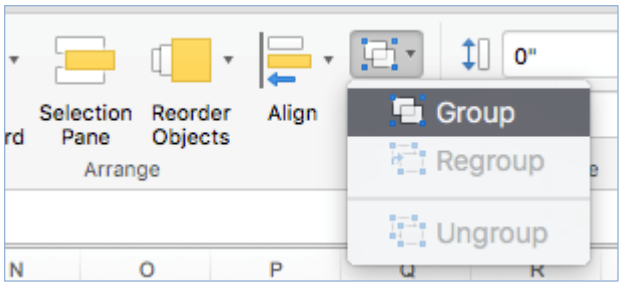
You may also click any of the axis label textboxes to change the displayed text and/or add text, or move the boxes using the arrow keys. You may also change the font size and style.

Copying the output: If you want to copy the normal distribution output and paste it into another application such as Word or PowerPoint you need to group all of the objects into one object. You do this as follows:

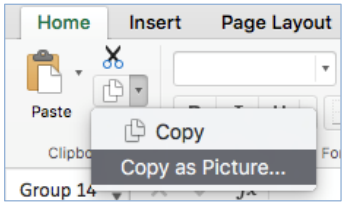
- Click anywhere on the graphical output to select one of the graphical objects.
- Click **Edit** → **Select All** on the menu bar at the top of the screen
- Click **Shape Format** → **Arrange** → **Group** → **Group** as shown below:



or



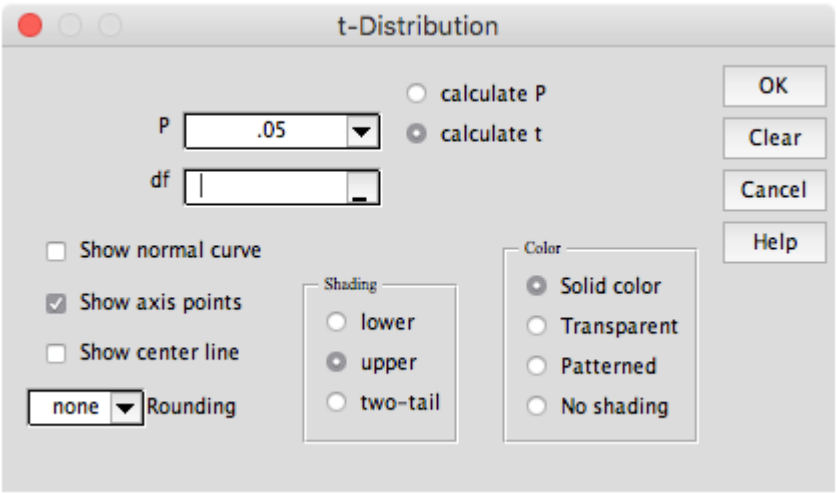
You can then right-click and **Save as Picture…** Or you can click **Home tab** → **Copy as Picture** and then paste it into other applications.



Display size: You can use the Excel zoom options make the display larger or smaller. You can also group all of the graphical options as shown in the step above and then drag a corner of the graphic to resize it.

Printing: To make the normal distribution output fit on a printed page without resizing it, you can use Page Setup or Print Preview and select Scale to fit.

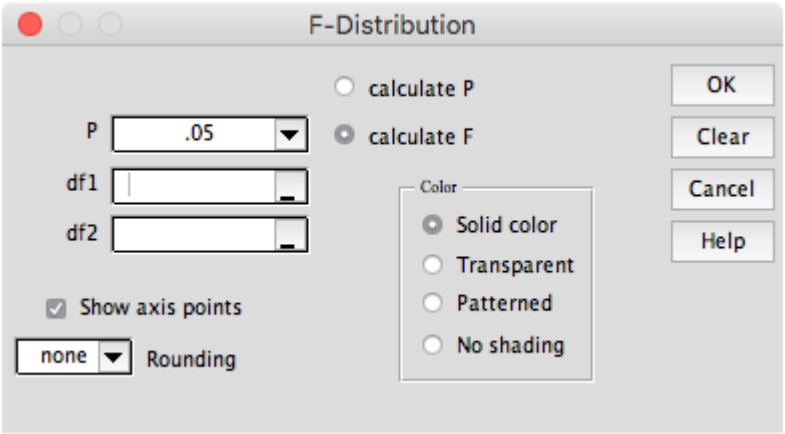
t Distribution



The general operating procedures for the t, F and chi-square distributions are the same as the normal distribution; however, the default option for them is to calculate the distribution value given a probability since they would often be used to determine critical values for hypothesis tests. See Tutorial Example 2 and the normal distribution reference section. The MegaStat help system gives detail regarding specific options for the t distribution.

Note in particular there is an option to superimpose a dashed line normal curve on the t distribution. For degrees of freedom of 30 and larger, the normal and t distributions will be nearly identical.

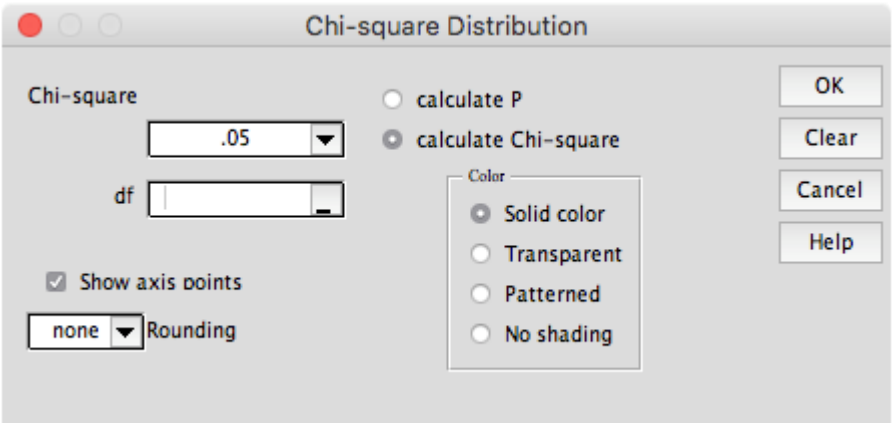
F Distribution



The general operating procedures for the t, F and chi-square distributions are the same as the normal distribution. See Tutorial Example 2 and the normal distribution reference section. The MegaStat help system gives detail regarding specific options for the F distribution.

The size of the graphical display is scaled to fit the display area and the maximum F value for the horizontal axis is chosen to be slightly larger than the .01 critical value. Thus the absolute size of displays for different degrees of freedom cannot be compared.

Chi-square Distribution



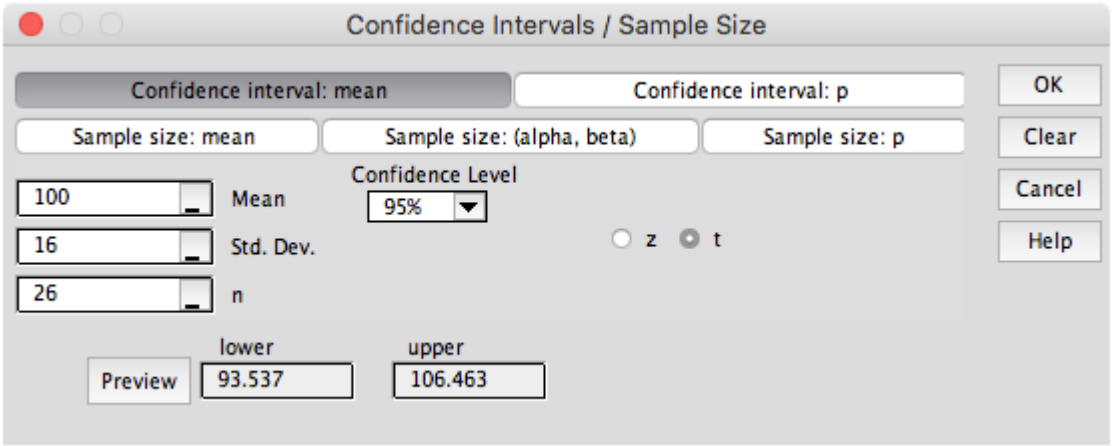
The general operating procedures for the t, F and chi-square distributions are the same as the normal distribution. See Tutorial Example 2 and the normal distribution reference section. The MegaStat help system gives detail regarding specific options for the chi-square distribution.

The size of the graphical display is scaled to fit the display area and the maximum chi-square value for the horizontal axis is chosen to be slightly larger than the .01 critical value. Thus the absolute size of displays for different degrees of freedom cannot be compared.

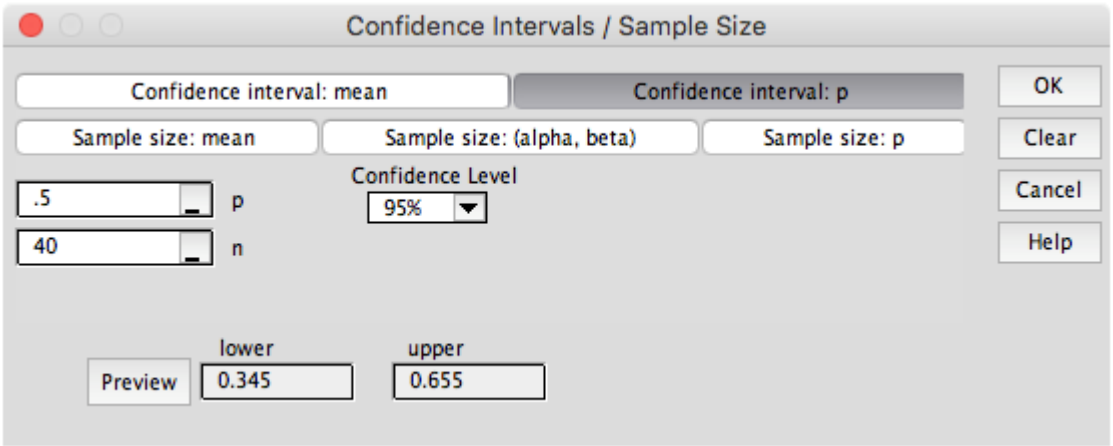
Confidence Intervals / Sample Size

For each of the dialog boxes confidence levels of .99, .95, or .90 may be selected by the dropdown arrow or you may type in any other value.

Confidence interval: mean

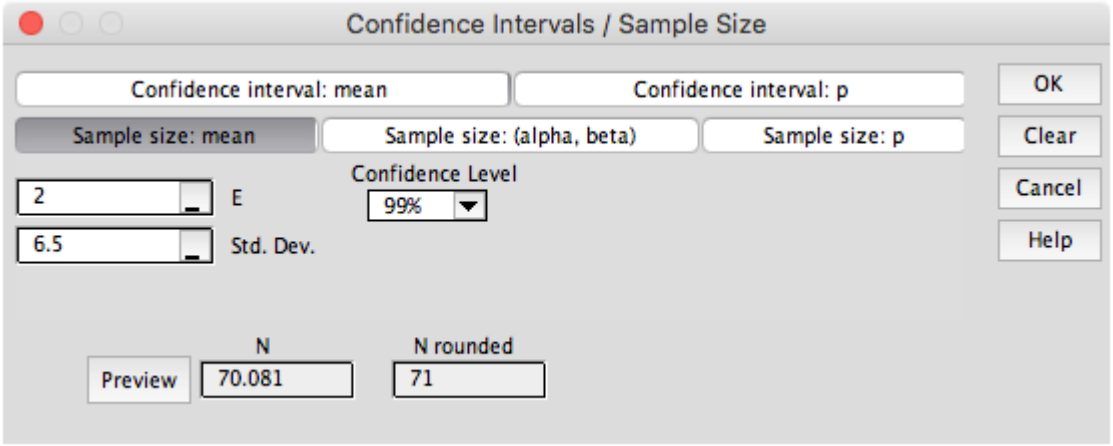


Confidence interval: p

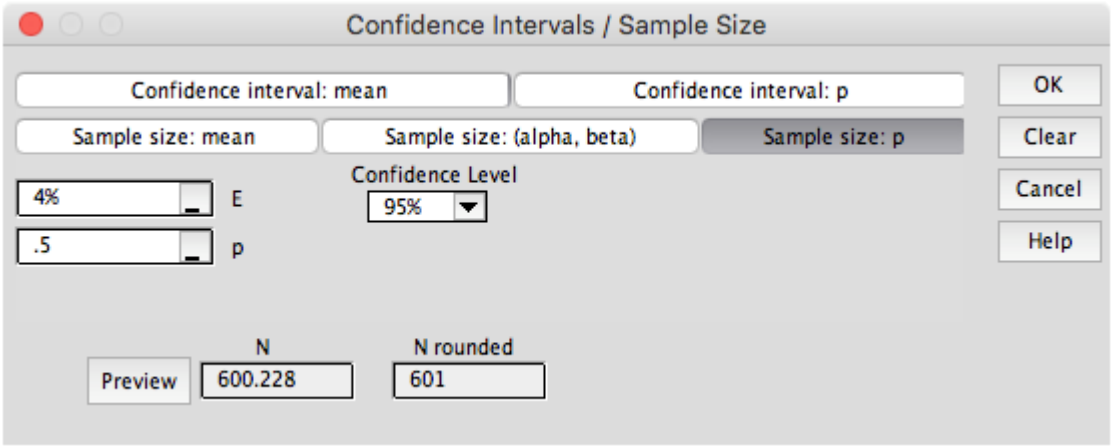


Note: Instead of typing .5 you could have typed 20 and the label would have automatically changed to 'x'

Sample size: mean

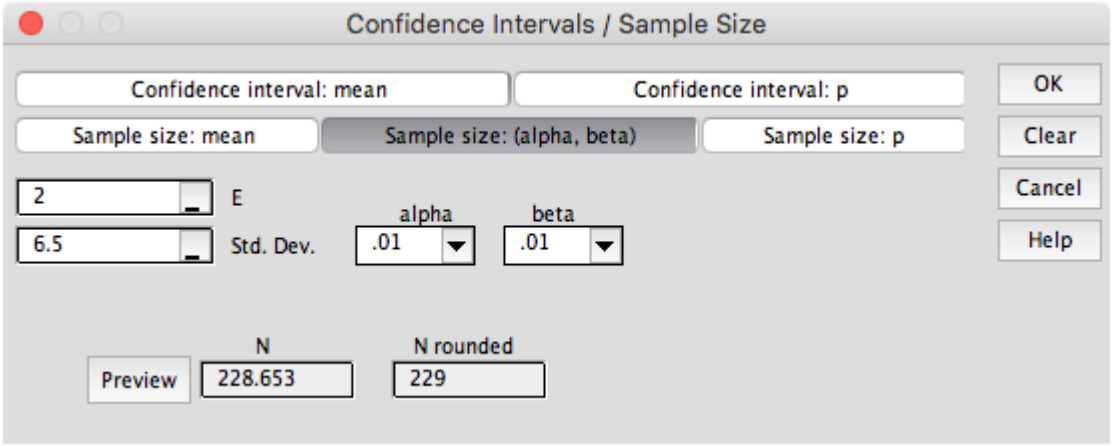


Sample size: p



Note that the Error tolerance could also have been entered as .04.

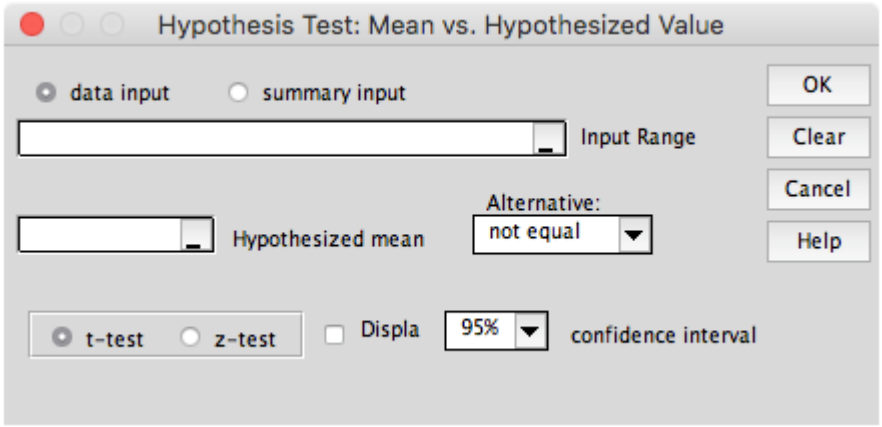
Sample size: mean with specified α and β



Note that sample sizes are rounded up to the next highest integer.

Hypothesis Tests

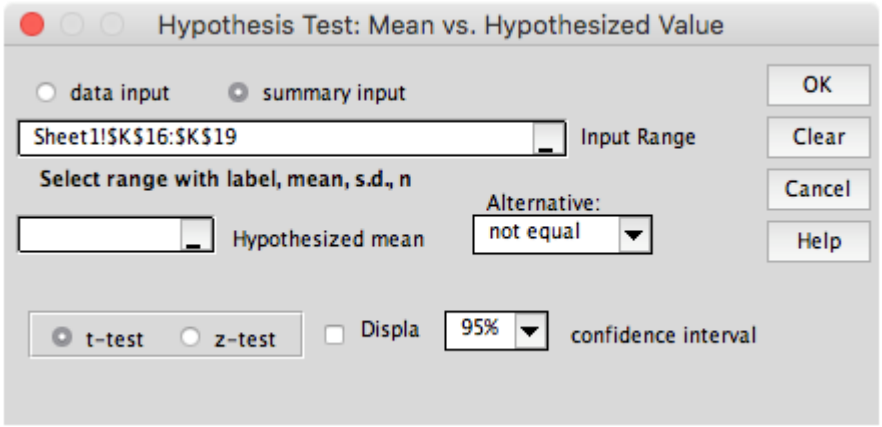
Mean vs. Hypothesized Values



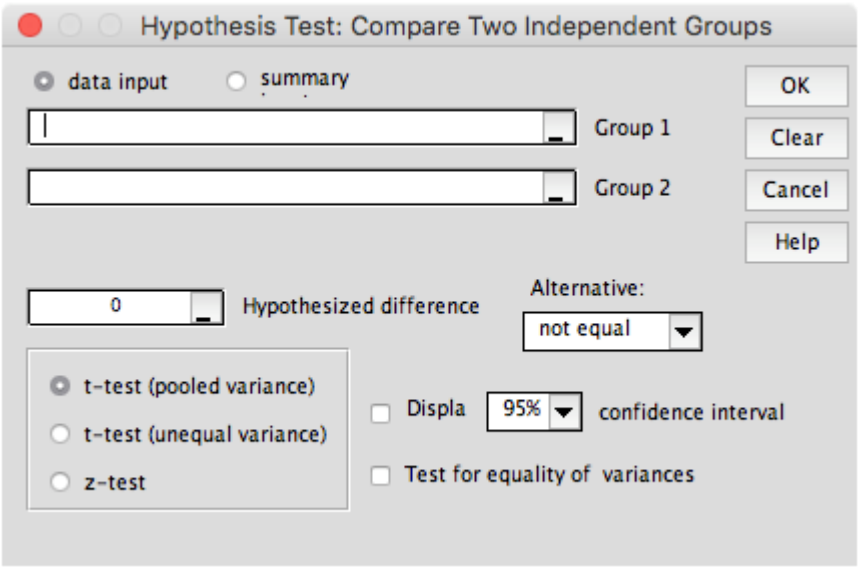
The t-z option determines what distribution is used for calculating the p-values. In either case the calculation is done using the sample standard deviation formula. If you want to use the population standard deviation formula, you would use the Descriptive Statistics option to do the calculation and then use the summary input option described in the next paragraph.

Summary input: Hypothesis test are usually done by specifying data input; however, there may be situations where you already have calculated values or want to specify known values for the standard deviation. For example, if cells K16 through K19 contain a label, mean, standard deviation and n respectively, you would click 'summary input' and select the cells as shown below. (If there is not a label, click a blank cell.)

If the summary values are not in contiguous cells you may select any four cells in the proper sequence while holding the **command** ⌘ key. This is useful if you are selecting summary data from Descriptive Statistics output.

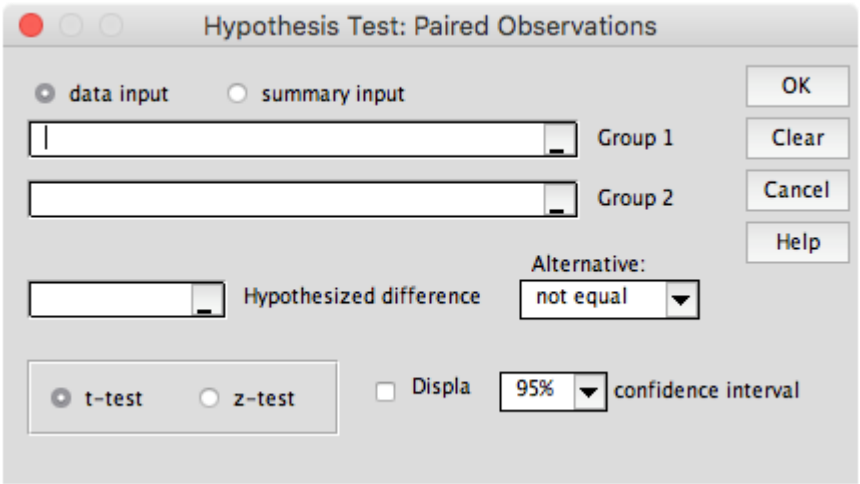


Compare Two Independent Groups



All of the data selected for each group will be treated as a single group even if you select multiple columns. (See the Mean vs. Hypothesized test for details on how to use the summary input option.)

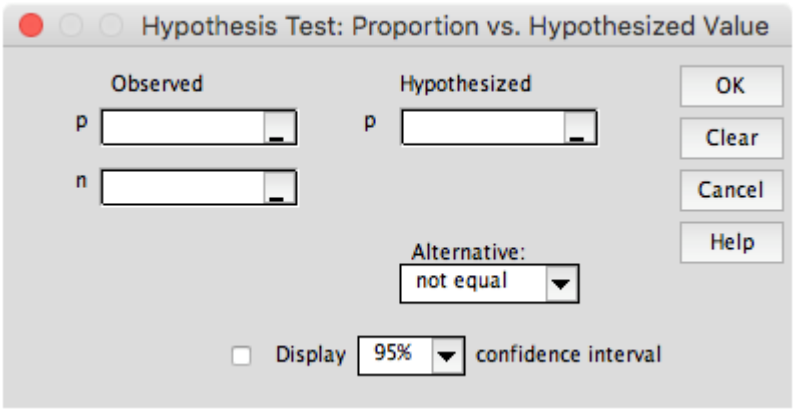
Paired Observations



Both groups must have the same number of observations. If your data is already in the form of differences, you would use the Mean vs. Hypothesized Value test.

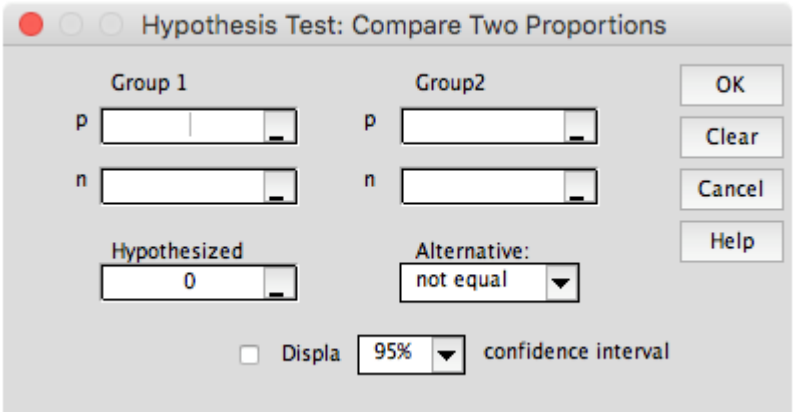
The t-z option determines what distribution is used for calculating the p-values. In either case the calculation is done using the sample standard deviation formula. If you want to use the population standard deviation formula, you would use the Descriptive Statistics option to do the calculation using the summary input option described in the Mean vs. Hypothesized section.

Proportion vs. Hypothesized Value

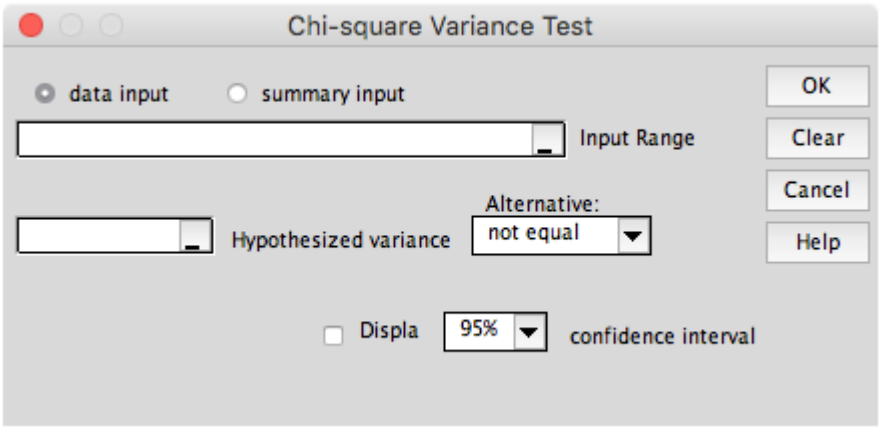


For this test and comparing two proportions, review the Entering Proportions section of the Tutorial Examples chapter.

Compare Two Independent Proportions



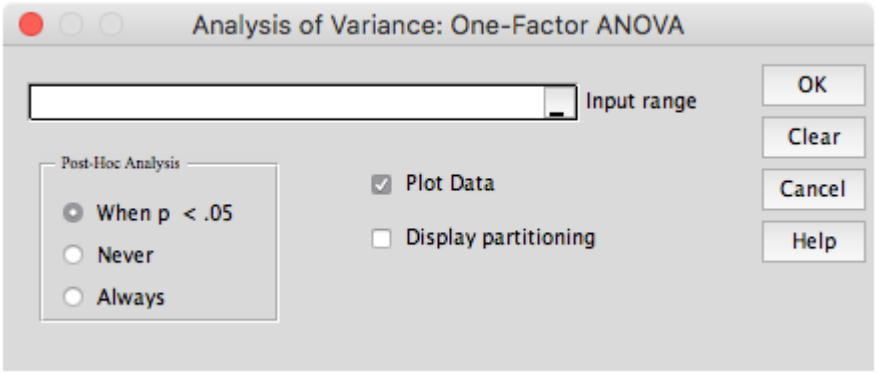
Chi-square Variance Test



To compare the variances from two groups use the hypothesis test for comparing two means and check the variance test option.

Analysis of Variance

One-Factor ANOVA



Within the input range each column will be considered a group. The groups do not have to be the same size so there may be some empty cells – just make the selection block large enough to include the largest group.

The data groups for this procedure must be side by side. For example, if you were doing an ANOVA with three groups the groups might look like this:

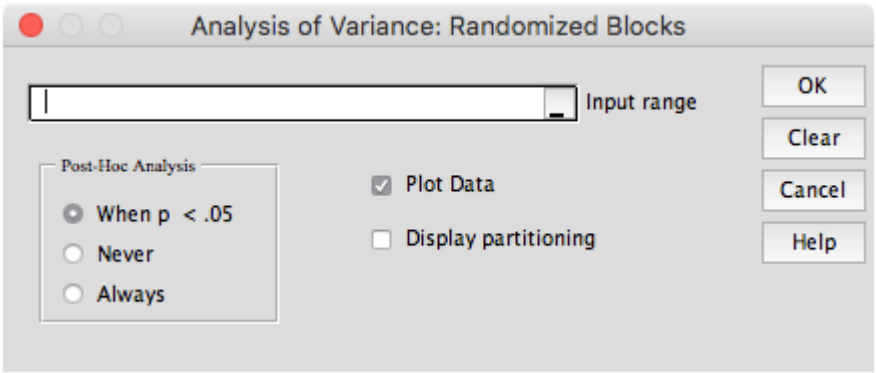
	A	B	C	D	E
1					
2		Group 1	Group 2	Group 3	
3		23	32	37	
4		25	29	31	
5		24	26	29	
6			33	31	
7			27		
8					

The post-hoc analysis is a table showing the p-values for the pairwise independent groups t-tests. The default option is to display the post-hoc analysis only when the ANOVA is significant at $p < .05$; however, you can specify that you never want the output or that you always want it.

Checking the Plot Data box gives a plot of the data with the group means and grand mean.

All three ANOVA's have a box to check if you wish to display partitioning. This will create output showing how the sums of squares are calculated.

Randomized Blocks ANOVA



Within the input range each column will be considered a treatment and each row will be a block. No missing or invalid data are allowed.

Each row (block) may have a label as well as the columns (treatments). Just remember that labels must be text. If you want to use AutoFill to generate block labels, enter the first two values as '1 and '2 to force them to be text.

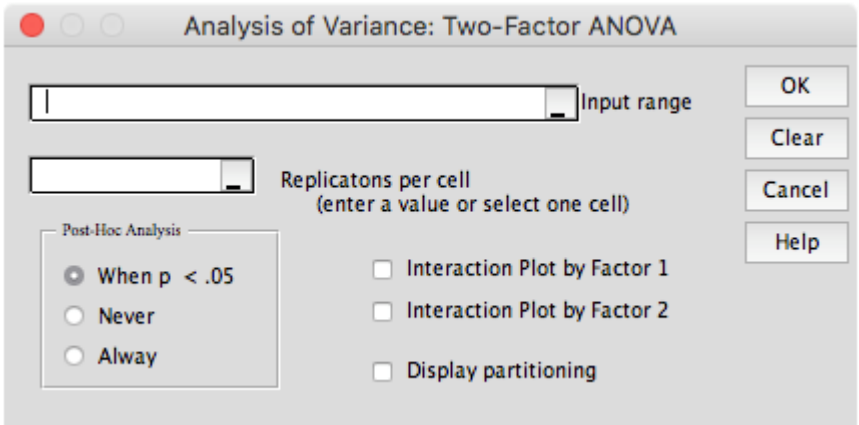
The data groups for this procedure must be side by side. For example, if you were doing an ANOVA with four treatments and five blocks the data cells must look like this:

	A	B	C	D	E	F
1						
2			Treatment 1	Treatment 2	Treatment 3	Treatment 4
3		Block 1	23	23	32	37
4		Block 2	25	26	29	31
5		Block 3	24	24	26	29
6		Block 4	27	25	31	34
7		Block 5	28	22	24	33
8						

The post-hoc analysis is a table showing the p-values for the pairwise independent groups t-tests. The default option is to display the post-hoc analysis only when the ANOVA is significant at $p < .05$; however, you can specify that you never want the output or that you always want it.

Checking the Plot Data box gives a plot of the data with the group means and grand mean. The data points for each block are connected.

Two Factor ANOVA



The data must be in the form shown below. Assume you have two column treatments, three row treatments and three replications per cell:

	A	B	C	D	E
1					
2			Column A	Column B	
3			48.2	55.7	
4		Row 1	53.7	52.5	
5			55.8	59.9	
6			73.0	76.2	
7		Row 2	78.1	78.4	
8			75.4	82.1	
9			52.4	54.0	
10		Row 3	49.7	52.1	
11			50.9	49.9	
12					

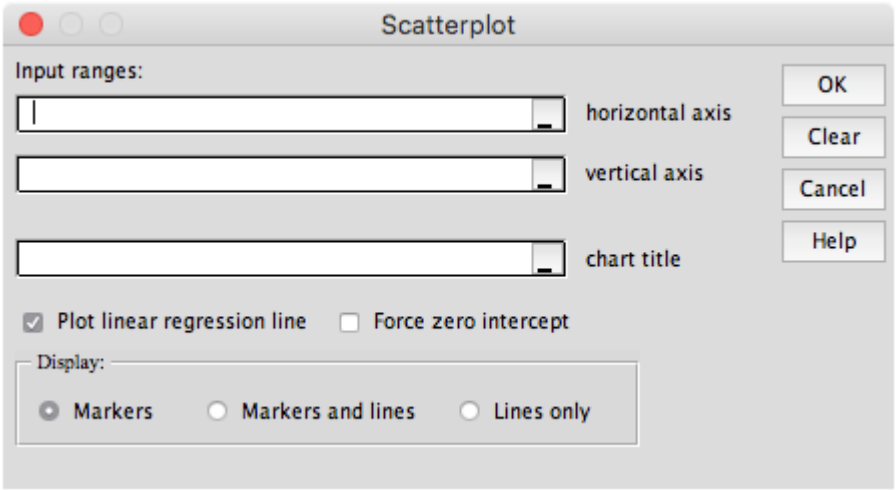
All cells must have the same number of replications and no missing data are allowed. You may type in the number of replications per cell or you can use the mouse to select one cell and the program will count the number of replications.

The row treatment labels may be placed in any of the cells to the left of the treatment.

The post-hoc analysis is a table showing the p-values for the pairwise independent groups t-tests. The default option is to display the post-hoc analysis only when the ANOVA for a given factor is significant at $p < .05$; however, you can specify that you never want the output or that you always want it.

Correlation / Regression

Scatterplot



Both input ranges must be the same size and be in a single column.

A title is not required but it is recommended. The title can have more than one line. Use the scroll bar on the edit box to view multiple lines. You may type in the title or click on a cell(s) that contains title text.

Specify if you want the linear regression line forced through the intercept.

Post output editing: MegaStat will do a good first approximation of the scatterplot, but remember that it is an Excel chart and if you right-click a chart object (*e.g.,* an axis) you go into editing mode where you can make changes. Some common things you might want to fine-tune are:

- Move the text box that contains the regression equation and r^2 if it is covering data points.

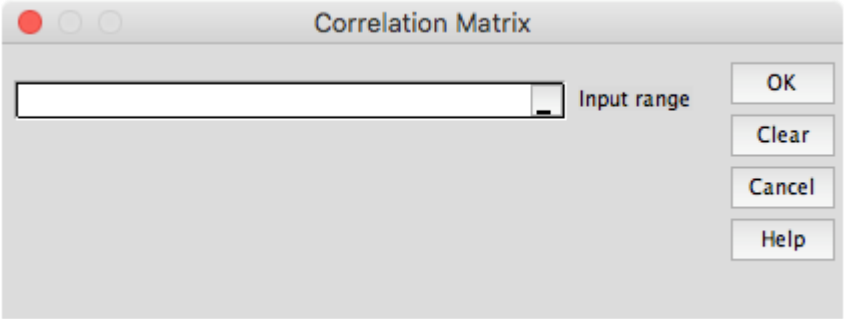
- Right-click the regression line (Trendline) and select Format Trendline → Options to extend the line forward or backward or try a non-linear curve fit.

- Rescale the axes. The program attempts to scale the axes properly; however, you can change the scaling to your preference.

- Resize the Scatterplot.

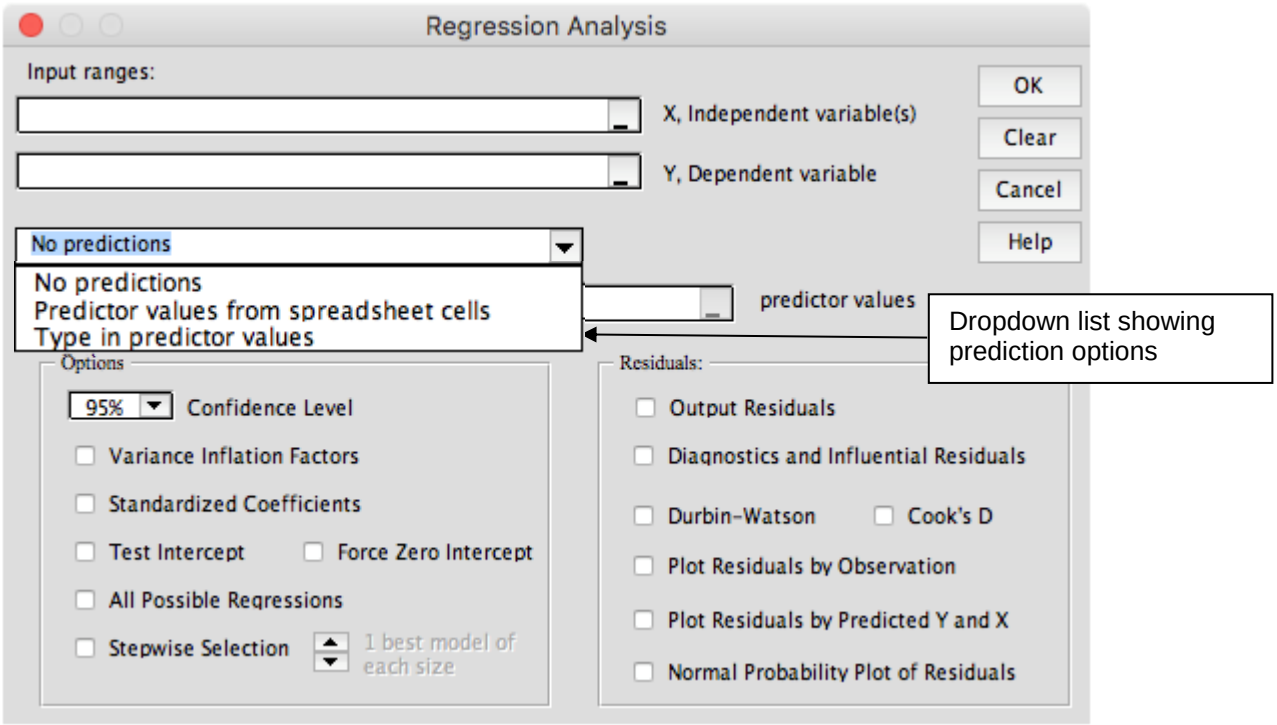
Note: The Scatterplot is linked to the data you selected in the input ranges. This means that if you subsequently change the data, the chart will be updated. If you want to lock it or paste it to another document without linking, do a Copy and Paste Special → Picture into Microsoft Word. It will look just the same, but you cannot do any further editing.

Correlation Matrix



If you want to change the order of the variables you may select multiple data ranges by holding the **command** ⌘ key. Each selection area must have the same number of rows. No missing or invalid data are allowed.

Only the lower half of the correlation table is displayed; however, the upper portion output is in hidden format and may be displayed by formatting the cells.



Within the input range each column is considered a variable. If there is more than one independent variable MegaStat will automatically do multiple regression. You may select multiple input ranges by holding the **command** ⌘ key. This is handy if you are doing multiple regression and the independent variables are not contiguous. The independent and dependent variable ranges must have the same number of rows.

If you want to calculate predicted values, click the dropdown arrow and select one of the options (as shown in the dialog box above):

Predictor values from worksheet cells

If you select this option you will need to select an input range for the values of the independent variable(s) to be used for prediction. Within this range each column will correspond to the independent variable(s) and each row will specify a different prediction. For example, if you are doing one prediction for simple regression the predictor range would be a single cell; or if you are doing three predictions for a multiple regression with four independent variables, the predictor range would be three rows by four columns.

The predictor range can be any cells, even in a different sheet; however, the obvious place to put the predictor variables would be right below the independent variables of the input data.

Type in predictor values

The input box allows you to type predictor values for the independent variable. The entries must be in the same order as the independent variables in the input data. The values must be separated by spaces or commas. You may make multiple predictions by typing a semi-colon between each prediction. If you have

several predictions and/or several independent variables it will be easier to enter them into worksheet cells and use the method described above.

For example: if you wanted to make three predictions using X1 = 20, X2 =9; X1 = 25, X2 = 12 and X1 = 30, X2 = 15 you would type in:

20,9;25,12;30,15

You can also enter spaces to make it easier to read

20,9; 25,12; 30,15

(Secret tip: You really don't need the semi-colons and commas if you put one or more spaces between each number. You could enter: 20 9 25 12 30 15)

Select Options:

-Confidence level for confidence intervals.

-Variance Inflation Factors to measure multicollinearity in multiple regression

-Standardized regression coefficients (sometimes called betas)

-Test for zero intercept

-Force zero intercept

-All Possible Regressions

This option prints a summary line for each possible combination of variables in a multiple regression. Since each summary line includes the p-values for the variables in the model, this output is useful for model building since you can easily see which combinations of variables work best.

The number of summary lines is $2^k - 1$ where k is the number of independent variables. Thus 12 variables will generate 4095 regression summaries. Above 10 variables the calculations can take quite a while and there may be memory issues. Stepwise Selection is usually a better option for more than 8 variables. The program still has to calculate the regression summaries but there is a lot less output since only the best regressions are displayed.

-Stepwise Selection

This option is similar All Possible Regression but it displays only the best models for any given number of variables, which gives more compact output.

All Possible Regressions and/or Stepwise Selection are more powerful and flexible than traditional Stepwise Regression. Stepwise Regression determines the one best model where MegaStat not only does this but also shows if there are other models nearly as good.

Selecting All Possible Regressions or Stepwise Selection causes all other options to be deactivated (grayed out). To restore the other options, uncheck All Possible Regressions or Stepwise Selection.

The Force zero intercept option will not work if the slope is negative or the intercept is not close to zero.

You cannot select Test intercept = 0 and Force zero intercept simultaneously. Selecting one will uncheck the other one.

Select Residuals options

- Output of dependent variable, predicted value, residual

  This option will automatically be selected if you choose any other residuals option.

- Durbin-Watson
- Plot Residuals by Observation
- Plot Residuals by Predicted Values and by each Independent Variable
- Diagnostic and Influential Residuals

  These are advanced options that would be important for upper level courses and research use.

  - --Leverage

    Shaded blue if the value is > 2 * (Nvar + 1) / n
  - --Studentized Residual
  - --Studentized Deleted Residual

    These are t values and they are shaded light blue if the p-value is <=.05 and blue if the p-value is <= .01
  - Cooks D

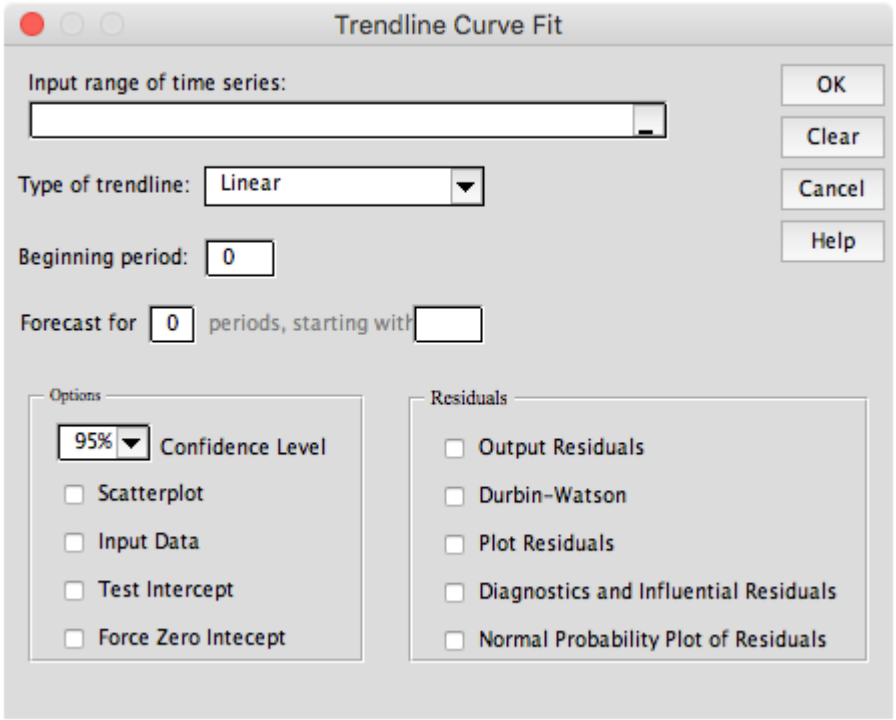
    This value is an F ratio and it is shaded light blue if the p-value is <= .80 and blue if the p-value is <=.50.

- Normal Probability Plot of Residuals

  Note: The data values needed to display the Normal Probability Plot are stored in a ChartDataSheet as discussed in the Utilities section.

Time Series / Forecasting

Trendline Curve Fit



Select a single column of data as the input range of the data to be fit with a trendline (*i.e.*, the dependent variable, Y). The program will create the time series independent variable (X). Then select the type of trendline you wish to fit: Linear, Exponential or Polynomial using the dropdown selection menu.

Select the beginning period for the time series variable (X). This is usually 0 or 1 but it can be any value, for example, a year. If you enter a non-integer value it will be rounded to the nearest integer.

If you wish to make one or more forecasts, specify how many and the starting period for the forecasts.

Specify if you want any of the output options:

Scatterplot of the data with the trendline. (In order to get the scatterplot you must also have the input data displayed so that box will automatically be checked also.

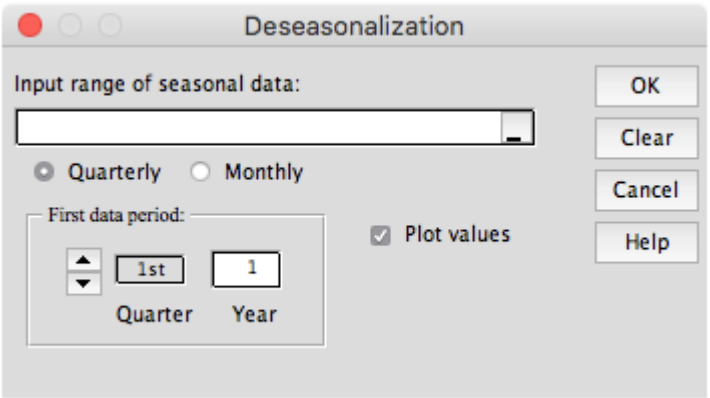
[In rare instances when you specify an exponential curve fit, the scatterplot will not be able to display the trendline. This is caused when you use large values (*e.g.*, years) for your time series. If you start the time series at t=0 the trendline will probably be displayed.]

Display the input data, *i.e.* the time series variable and the input data.

Output of residuals, Durbin-Watson and plot of residuals. Note that you cannot get the latter two unless you output the residuals.

Options for Adjusted R^2, Test for intercept and Force intercept through zero. (See regression section for details on these items.)

Deseasonalization

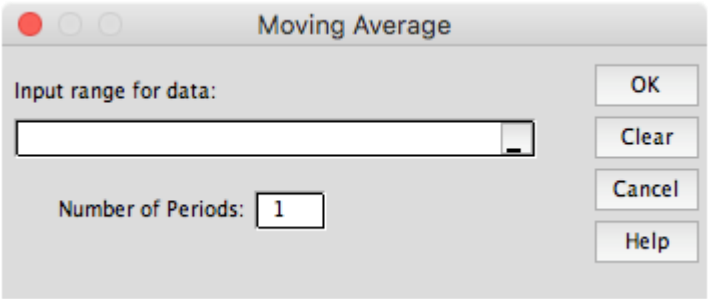


In order to do deseasonalization you must have at least two years of data (*i.e.*, 8 quarters or 24 months). You cannot have missing data; however, you can have a partial year at the beginning and/or end of the data.

Indicate if the data is quarterly or monthly. Then specify the quarter/month and year of the first data value. These values are used for labeling the output and are not required.

After you have run the analysis you can use the Trendline or Regression options to make predictions using the deseasonalized data. (Make sure to remember to reseasonalize the prediction by multiplying by the appropriate seasonal index.)

Moving Average

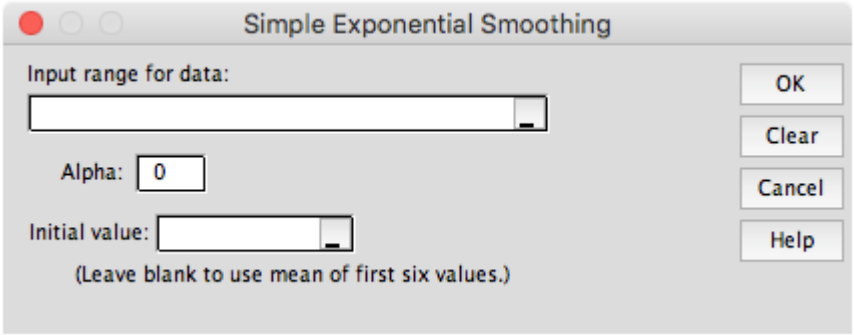


Select a single column of data as the input range. Then specify the number of periods in the moving average (usually 3 to 5 but can be any value between 1 and n). The example above shows a 5 term moving average.

The moving average is displayed adjacent to the last period averaged; however, you can cut/paste or drag/drop the averaged values to position them anywhere.

Exponential Smoothing

Simple Exponential Smoothing

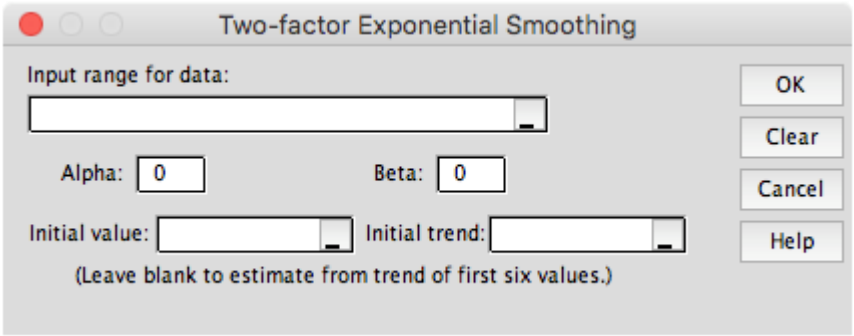


Select a single column of data as the input range of the data to be smoothed.

Specify Alpha, the weight to be given to each new value of the smoothed series. Alpha must be between 0 and 1.

Specify an initial data value for the smoothed series. If you leave the initial value blank, it will use the mean of the first six data values.

Two-factor Exponential Smoothing

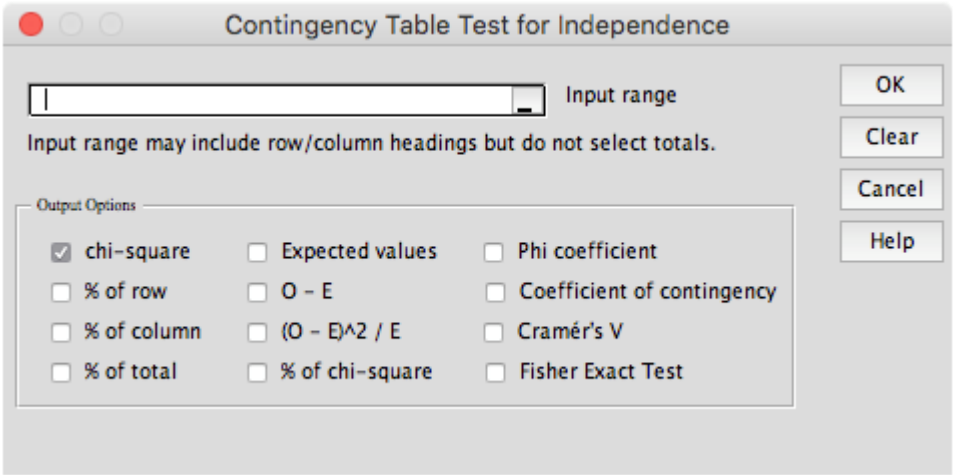


Select a single column of data as the input range of the data to be smoothed.

Specify Alpha, the weight to be given to each new value of the smoothed series. Alpha must be between 0 and 1. If you leave this blank it will use the <u>intercept</u> of the linear trend of the first six values. Then Specify Beta, the weight to be given to the trend. Beta must be between 0 and 1. If you leave this blank it will use the <u>slope</u> of the linear trend of the first six values.

Chi-Square / Crosstab

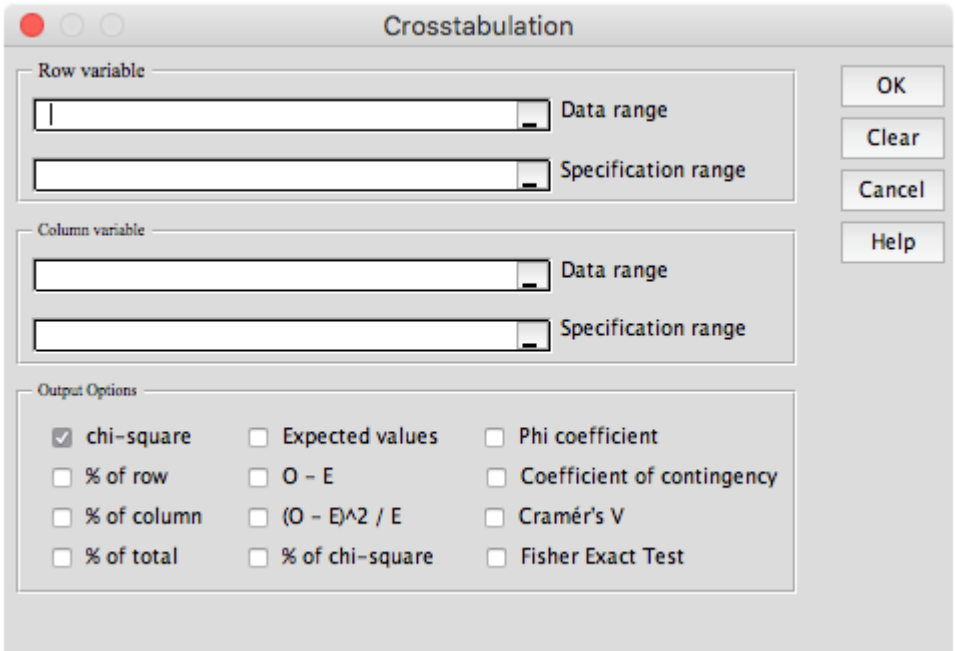
Contingency Table



You may include row and column headings in the input range; however, you should not include row or columns totals. The table cannot contain any empty or non-numeric cells

Click on the Output Options checkboxes to select the values to be output. In general, it is not a good idea to select all of the options for one output since this will give a cluttered output.

Crosstabulation

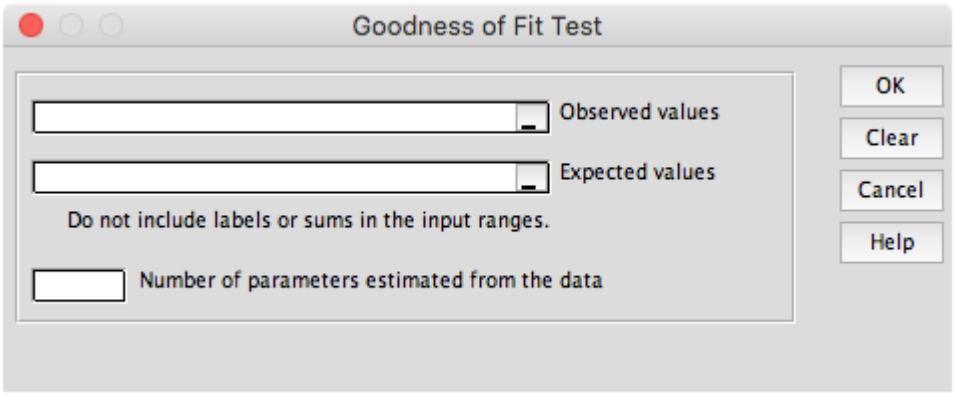


This option calculates a two-factor crosstabulation table from qualitative data.

Select the input data ranges for the row and column variables. Then select the specification range for the row and column variables. (See the discussion of specification ranges under Qualitative Frequency Distributions.)

Click on the Output Options checkboxes to select the values to be output. If all you want are the counts, leave all options unchecked.

Goodness of Fit Test

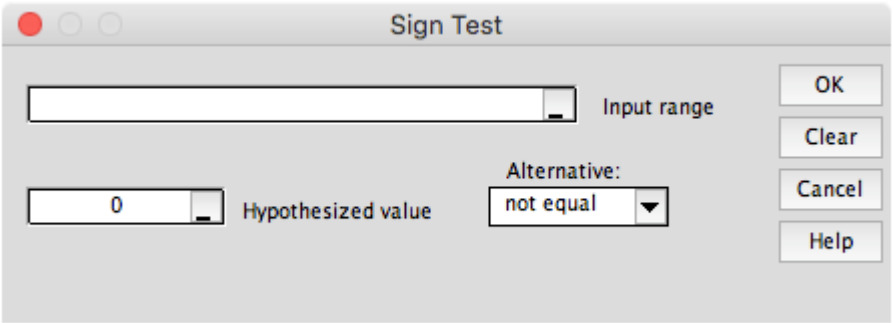


Select the input data ranges for the observed and expected frequencies. You must not include labels in the range. Both ranges must include the same number of cells although they don't have to be the same shape; however, if the ranges are not the same shape, carefully check the output to make sure observed and expected values were paired properly.

Specify the number of input parameters estimated from the data. The default value is none. For example, if you were doing a Goodness of Fit for a normal distribution and you used the mean and standard deviation from the actual data to find the expected values, then this value would be 2.

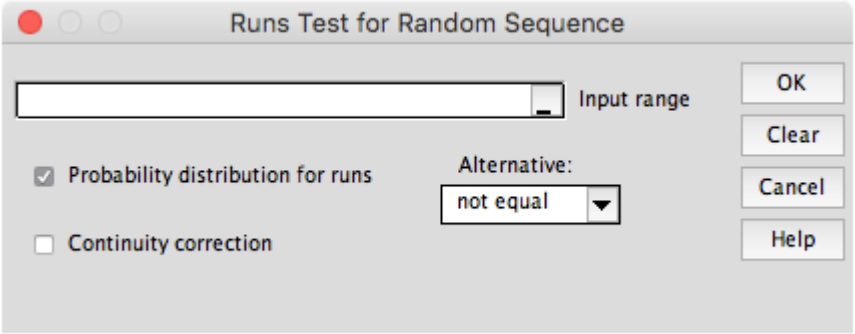
Nonparametric Tests

Sign Test



This test can also be performed with the binomial distribution with n = sample size and p = .5.

Runs Test for Random Sequence



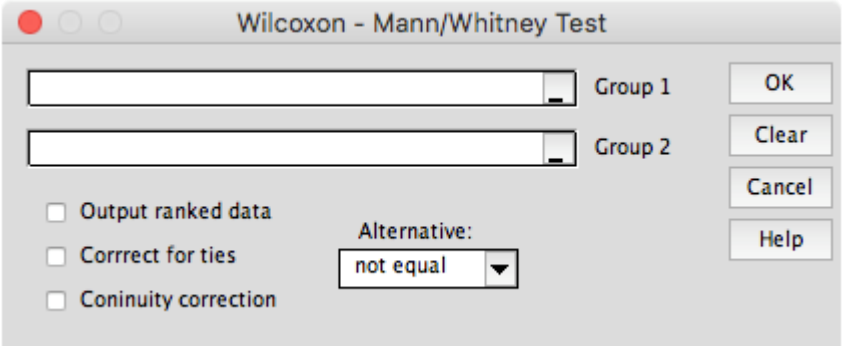
The program expects each cell in the input range to contain one of two unique values and it tests whether there are too few or too many runs than would be expected by chance. The two values can be either alpha or numeric but if more than two different values are detected in the range an error message will be displayed.

If you want to use this test to check runs above and below the median, use the Excel IF function to code your data into two values representing above and below the median.

Note: There is checkbox for a continuity correction; the default is off.

Check if you want to display the complete probability distribution for the total number of runs. Hint: If you want to see the probability distribution and do not have a set of data, create a dummy dataset with n1 A's and n2 B's (or any two values).

Wilcoxon – Mann/Whitney Test

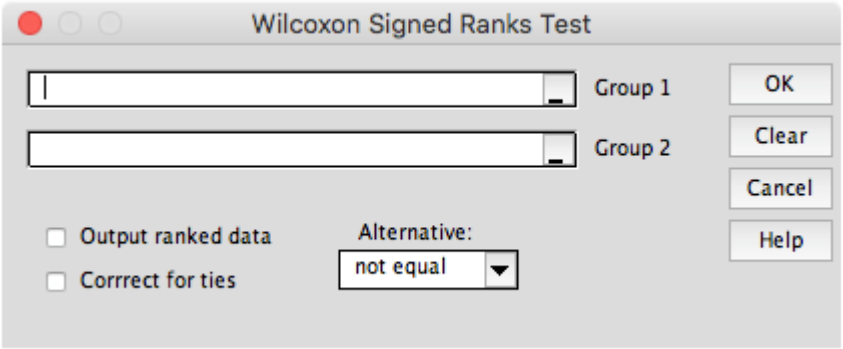


This test works with ranked data; however, the data does not have to be ranked. The program will convert the data to ranks as required by the test.

If you check that you want to output ranked data, the ranked data will be on the output sheet; your original data will not be changed.

Check if you want to correct for ties and/or do a continuity correction.

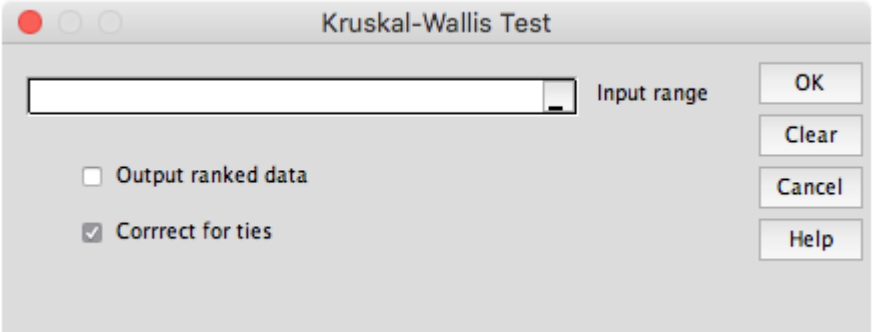
Wilcoxon Signed Ranks Test



This test works with ranked data; however, the data does not have to be ranks. The program will convert the data to ranks as required by the test. Since this test works with paired observations, both groups must be the same size.

If you check that you want to output ranked data, the ranked data will be on the output sheet; your original data will not be changed. Check if you want to correct for ties.

Kruskal – Wallis Test



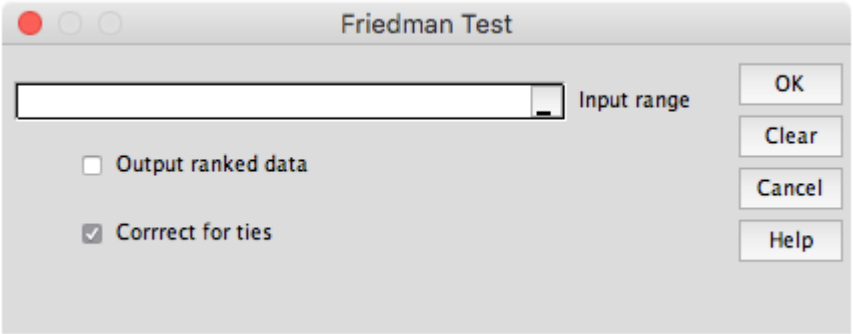
Within the input range each column will be considered a group. The groups do not have to be the same size so there may be some empty cells – just make the selection block large enough to include the largest group. See One Factor ANOVA for details on data layout.

This test works with ranked data; however, the data does not have to be ranks. The program will convert the data to ranks as required by the test.

If you check that you want to output ranked data, the ranked data will be on the output sheet; your original data will not be changed.

The default is to correct for ties and generally should be selected.

Friedman Test



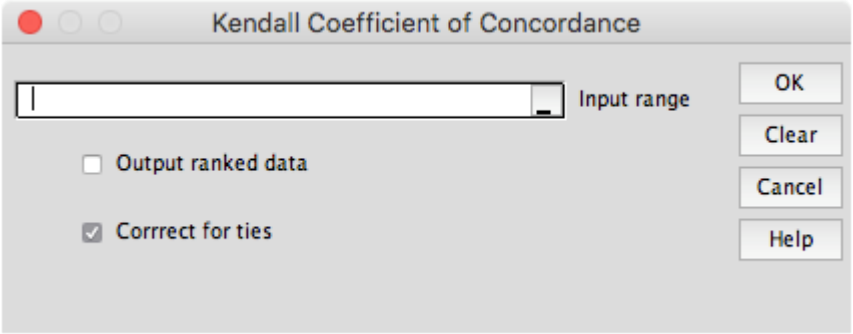
Within the input range each column will be considered a treatment and each row will be a block. No missing or invalid data are allowed. See Randomized Blocks ANOVA for details on data layout.

This test works with ranked data; however, the data does not have to be ranks. The program will convert the data to ranks as required by the test.

If you check that you want to output ranked data, the ranked data will be on the output sheet; your original data will not be changed.

The default is to correct for ties and generally should be selected.

Kendall Coefficient of Concordance

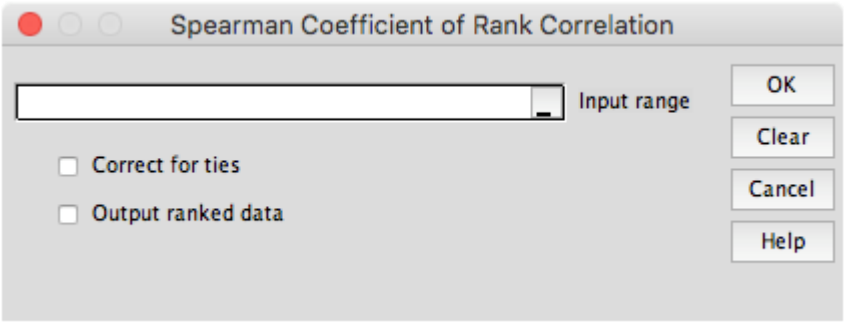


Within the input range each column represents an item being judged and each row represents a judge. No missing or invalid data are allowed.

This test works with ranked data; however, the data does not have to be ranks. The program will convert the data to ranks as required by the test.

If you check that you want to output ranked data, the ranked data will be on the output sheet; your original data will not be changed.

Spearman Coefficient of Rank Correlation

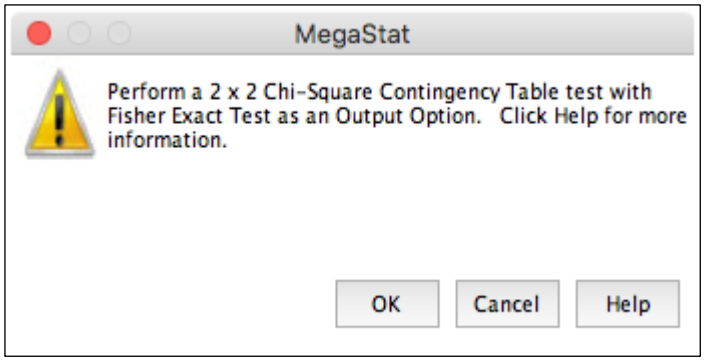


Within the input range each column will be considered a variable. No missing or invalid data are allowed.

This test works with ranked data; however, the data does not have to be ranks. The program will convert the data to ranks as required by the test. The program can correct for tied ranks.

Check whether you want to correct for ties and check whether you want to output the ranked data. The ranked data will be on the output sheet; your original data will not be changed.

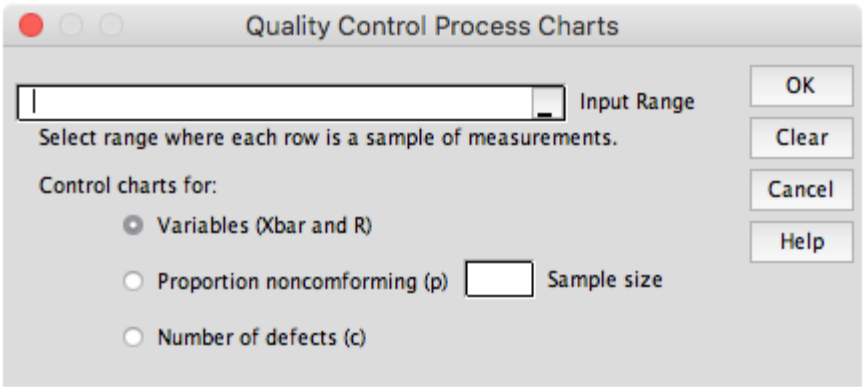
Fisher Exact Test



This test is an output option of the chi-square Contingency Table Test and assumes you have data in the form of a 2 x 2 contingency table. When you click OK the Contingency Table Test will be loaded with the Fisher Exact Test checked as an output option. You may also select other output options.

Quality Control Process Charts

This option allows you to plot the most common quality control process charts.



Control chart for variables (Xbar and R chart)

Select a range of data where each row is a sample of measurements. You may have missing data in any sample; however, the sample size for the chart will be determined by the number of columns in the selected range.

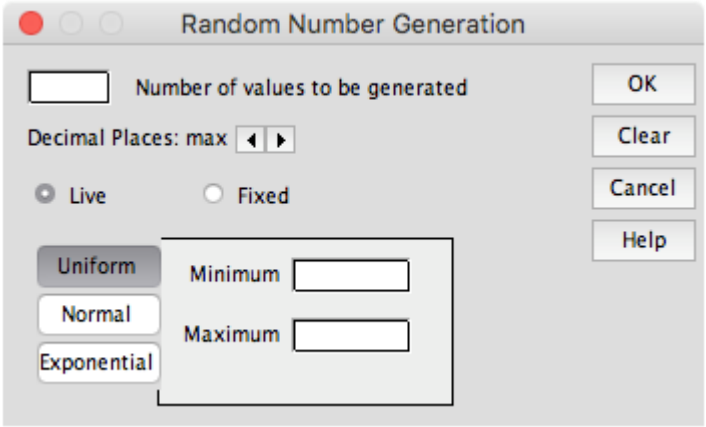
Control chart for proportion nonconforming (p chart)

Specify a sample size and select a range where the values are the proportion or number of defective items in each sample. If the number is less than one it will be considered a proportion; if the number is greater than or equal to one it will be divided by the specified sample size to calculate the proportion.

Control chart for number of defects per sample (c chart)

Select a range where each value is the number of defects in a sample.

Generate Random Numbers



If you specify live functions MegaStat places Excel functions in the specified number of cells. You may recalculate the values by pressing the control-= or F9 function key (fn-F9 on some keyboards). If you find a set of values you want to keep, you can 'freeze' them by selecting the cells and doing a Copy followed by a Paste Special → Values.

If you select uniform random numbers you need to specify the minimum and maximum values; the normal distribution will need a mean and standard deviation; and the exponentially distributed numbers require, mu, the mean rate of occurrence.

The output calculates the mean, standard deviation, n, minimum, and maximum of the random numbers. The standard deviation is calculated using the sample formula STDEV.S. If you want the population formula, change the cell formula to STDEV.P.

The exponential distribution creates a distribution skewed to the right. If you want a distribution skewed to the left, subtract the values from a constant larger than the largest value.

Appendix A. Frequently Asked Questions and Technical Information

This appendix deals with general issues related to MegaStat version 10.4 for Mac Excel.

1. **MegaStat is an add-in program. What is an add-in?**

   An Excel Add-in such as MegaStat is a program that provides extended functionality to Excel. Once MegaStat is installed it will be available on the Data ribbon whenever Excel is opened.

   The advantage of having MegaStat as an add-in is that Excel is used for data entry, data manipulation, and printing. That allows MegaStat to be a relatively compact program that focuses on statistics and without having to deal with data management. Also Excel is a very common data file format and most other file types can be converted to Excel files.

2. **When I try to open the MegaStat.pkg download I get a message saying it can't be opened because it is from an unidentified developer. What should I do?**

   The message is because the download is not from the Apple app store. Open the Downloads folder with Finder, right-click or control-click MegaStat.pkg, and click Open.

3. **How are the Mac versions of MegaStat different than the Windows versions?**

   The appearance is slightly different, but they are quite similar and have all the same options and output formats. The Windows and Excel versions use the same file format which means that files developed on one version can be used with the other version.

   The Windows versions uses a dedicated Help viewer and the Mac versions uses a web browser.

4. **Can MegaStat be moved to another ribbon or the Quick Access Toolbar?**

   This cannot be done with the current version (January 2019) of Mac Excel. It is possible that this feature will be implemented on future versions.

5. **What can I do if Excel 'locks up' on me?**

   Hopefully this won't happen to you but if something happens that makes Excel nonresponsive, hold down the **command-option-esc** keys to activate the Force Quit option.

6. **How large is MegaStat?**

   MegaStat.xlam and the help files are less than 2MB.

   MegaStat does not require a large amount of memory. If you can run Excel, you can probably run MegaStat**.**

7. **Does MegaStat have data size restrictions.**

Multiple regression is limited to 12 predictor variables; however, MegaStat does not generally have arbitrary restrictions. MegaStat (and Excel) will work best with moderate size data sets.

8. **Will MegaStat run on a network?**

MegaStat was not developed specifically for network use and is not supported for network use; however, some users have reported successful network usage.

Check with McGraw-Hill regarding network licensing before attempting a network installation.

9. **Where are the MegaStat files stored?**

MegaStat for Mac Excel is installed in the Applications/MegaStat folder.

10. **How does the MegaStat help system work?**

There are approximately 160 html files in the MegaStat Help folder. MegaStat uses the default browser (usually Safari) to display the files as needed. Apple Safari, Google Chrome, Microsoft Edge, and Firefox browsers have been tested; other browsers should work.

When displaying MegaStat help using Safari, and possibly other browsers, the browser focus may not be brought to the top so the help window will be partially or completely hidden by the MegaStat window. Click on the browser window or the browser icon in the Dock to bring the help window to the top. This does not happen with Google Chrome or Microsoft Edge and hopefully this issue will be fixed for other browsers with a future update.

Applications/MegaStat/MegaStat_Help/idh_MegaStatHelp.htm is the main help file and can be opened directly with your browser if you want to view the help system without running Excel and MegaStat.

11. **When I click Help in MegaStat nothing appears.**

MegaStat uses the default browser, usually Safari, to display help. Safari does not automatically come to the front when Help is clicked. If you see just part of the Help window, click it to bring it to the front. If you do not see the Help window, click Safari in the Dock.

12. **Can MegaStat be used without running the installer program?**

You need to initially run the installer (MegaStat.pkg) which will put the files in the Applications folder. However, you can copy MegaStat.xlam and the help file folder to any folder and then Browse to that folder from the Add-Ins list.

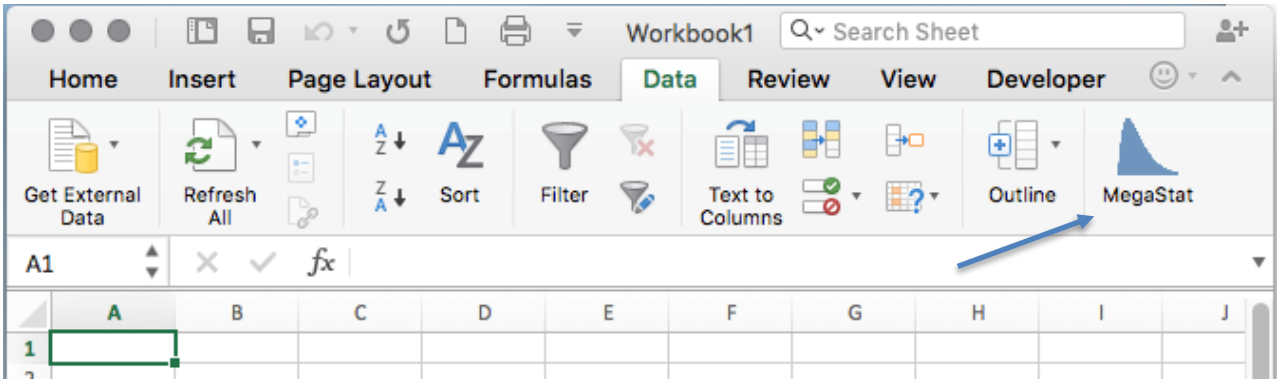
13. **I get errors when I enter decimal values, what is the problem?**

This is probably caused by changing the decimal separator from a period to a comma. MegaStat requires that the decimal separator be a period.

Appendix B. Installation Instructions

This version of MegaStat will only work with Mac Excel 2016, 2019, or Microsoft 365. It will not work with Mac Excel 2011, 2008, or earlier.

After the installation MegaStat will appear on Excel's Data ribbon as shown below.



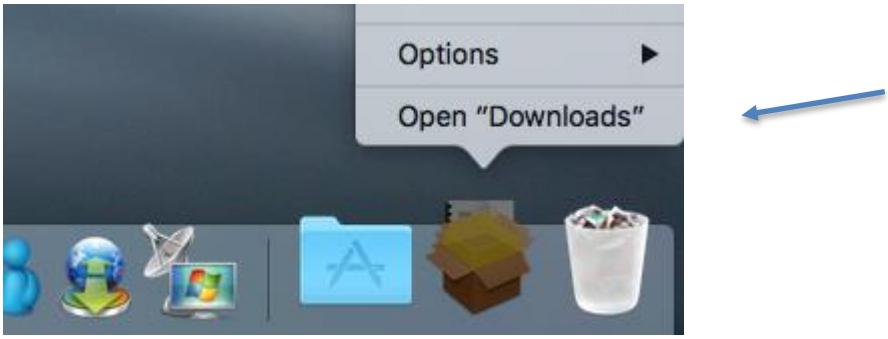
MegaStat for the Mac Excel has all of the features of the Windows version and works the same way. The design and appearance of some of the dialog boxes will be slightly different from the Windows version but the functionality is the same.

The MegaStat installation and setup is shown below. It is probably faster to do it than to read about it but the steps are shown in detail below.

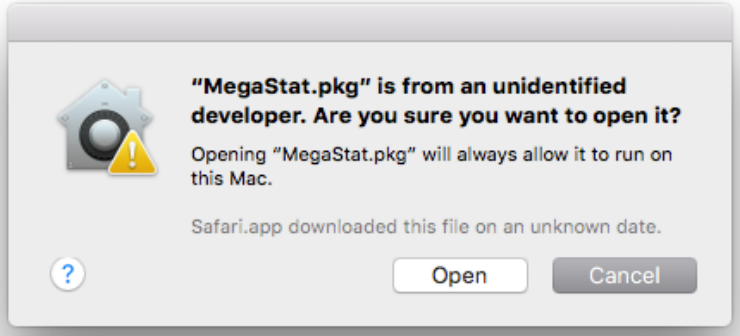
MegaStat installation

- Download the installation program (*MegaStatInstallerMac.pkg*). You will need Administrator privileges to do the installation.

- Open the Downloads folder in Finder by **control-clicking (or right-clicking)** the Downloads icon in the dock and clicking **Open Downloads**. (Or you can just open Finder from the dock.)



- In the Downloads folder, **control-click** or **right-click** *MegaStatInstallerMac.pkg*, click **Open**, and you will see the following:



(This message is because MegaStat is not coming directly from the Apple app store.)

- Click **Open** to start the installer program and view the Introduction.
- Click **Continue** to view the License Agreement. There are option buttons to print or save the file.
- Click **Continue** and click **Agree** to accept the License Agreement.
- Click **Install** to start the installation.
- You will be asked for your username and password. Then click **Install Software**.
- An Excel workbook will then be opened that will place MegaStat on the Data ribbon. By the time you see the workbook, the setup will have been completed. If you wish you can verify that MegaStat is on the Data ribbon. Close the workbook and the click Close to exit the installer.

  Whenever you open Excel, MegaStat will be available on the Data ribbon. (Note that MegaStat is on the Data *ribbon* and not the Data item on the menu bar at the top of the screen.)

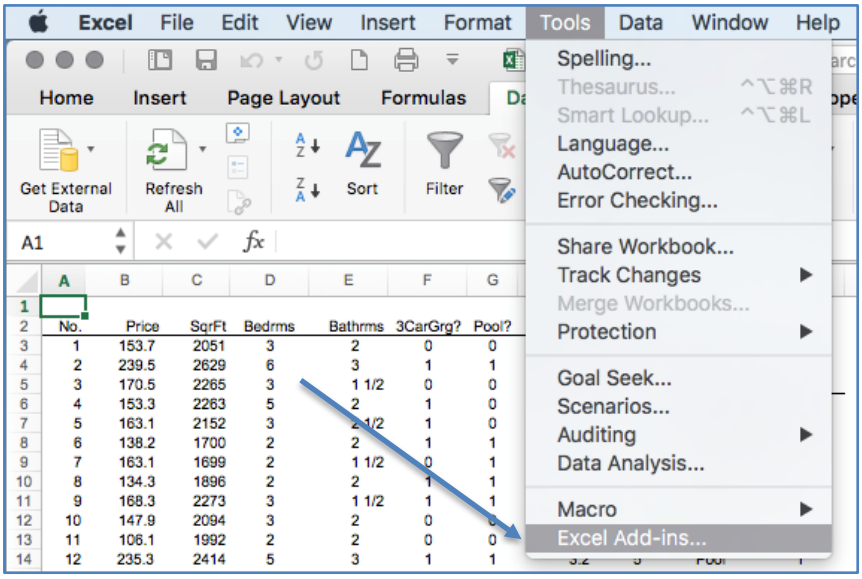
Manual MegaStat setup

You shouldn't have to do this but in case you want to know how add-ins work, here are manual setup steps. The installation workbook does these steps automatically when it opens.

The installation program copies the MegaStat files to the MegaStat folder in the Applications folder. Here are the steps in Excel to set MegaStat up as an add-in. MegaStat will then be available on the Data ribbon whenever you run Excel.

1. **Open Excel**. If there isn't a workbook open, start a new one or open any existing workbook.

2. Click **Tools > Excel Add-ins…** to open the Add-ins window.



3. In the Add-ins window click **Browse**… That should open the Applications folder. If not, find the Applications folder.

4. In the Applications folder click the **MegaStat folder**, click **MegaStat.xlam**, and then click **Open** which will take you back to the Add-ins window?

5. MegaStat will appear checked in the Add-ins list. Click **OK**.

6. MegaStat will now be at the right-hand end of the Data ribbon and will be there every time you open Excel. (Note that MegaStat is on the Data *ribbon* and not the Data item on the menu bar at the top of the screen.)

If you want to remove MegaStat from the Data ribbon, go to Tools -> Excel Add-Ins... and uncheck MegaStat.

To uninstall MegaStat, use Finder to remove the MegaStat folder from the Applications folder.