

Excel for Statistics

Office of Information Technology

West Virginia University

oithelp@mail.wvu.edu, (304) 293-4444 – Help Desk

<http://oit.wvu.edu/training/classmat/xl/>

Table of Contents

Data Entry in Excel.....	2
Suggestions and Tips	2
The Fill Handle	3
Formulas and Functions.....	4
Formulas.....	4
Order of Operations	4
Functions	4
AutoSum.....	6
Standardize Scores	7
Practice using statistical functions	7
About statistical analysis tools.....	7
Install Analysis ToolPak	8
Descriptive Statistics	9
Histogram.....	10
Sparklines	11
Insert a Sparkline.....	11
Customize the Sparkline.....	12
Chi-Square Test	13
T-Test.....	14
Anova: Single Factor.....	15
Standard Deviation Enhanced Line Graph	16
Import Excel Data into SPSS	17
Import Excel Data into SAS.....	18
References.....	18

Data Entry in Excel

It doesn't matter if you plan to use Excel solely as a way to enter your research data for future analysis in SAS or SPSS or if you intend to use Excel to analyze your data – you still need to get the information into Excel. This section assumes you are entering your own data as opposed to extracting it from another tool, purchasing a database, or downloading information from a web site.

	A	B	C	D	E
1	id	county	name	bday	score
2	1	17	Tropf	1/22/1902	47
3	2	17	Wright	3/30/1900	55
4	3	19	Mitchell	7/15/1962	22
5	4	19	Craig	4/2/1964	58

Suggestions and Tips

- Enter each observation in a new row; each variable will have its own column.
- Use row 1 to enter your labels (also known as column headings or variable names).
- Keep variables names meaningful and short with no blanks or special symbols.
- Start entering data in row 2.
- Do not have any completely blank rows or blank columns.
- If possible, make sure the first line of data contains the type of data you expect for the whole column (numeric vs character).
- If a data item is missing, leave the cell blank.
- As you enter data, pressing the tab key will move your cursor to the cell to the right. Press Enter to go to the next row.
- Use the mouse to click in the cell where you want to modify information or use the arrow keys to move one cell at a time.
- Ctrl Home will take you to cell A1. Ctrl End takes you to the bottom right corner.
- To delete data in a cell, click once on the cell and press the Delete key.
- To replace data, click once on the cell and type new information then press Tab or Enter.
- Verify you are in the correct cell before you enter a number; you can accidentally wipe out previously entered data (see above).
- To change information, click once on the cell to select it and then click on the text on the formula bar to edit it. -OR- double click on the cell and edit it in place.
- You can undo previous action(s) by choosing Undo from the Edit menu, by clicking on the Undo button (looks like a bent blue arrow), or by pressing **Ctrl Z**
- A shortcut to enter today's date: press **Ctrl** and ; keys at the same time
- A shortcut to duplicate information from the cell in the previous row: **Ctrl '**
- Since you are mainly interested in the data, do not waste time formatting at this time (alignment, fonts, shading, colors, styles, etc.).

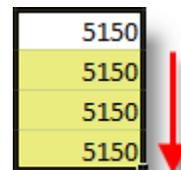
- Save early; save often (**Ctrl S** or click on button that looks like a diskette). Make backup copies of your data in more than one location (e.g., hard drive, USB drive, a local area network drive, email it to yourself, use DropBox, etc.).

The Fill Handle



Copying Information

An alternative to using Copy and Paste when you wish to copy information or formulas to adjacent cells is to take advantage of the AutoFill feature.

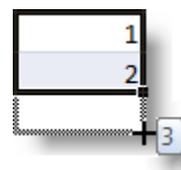


If you want to copy the contents of a cell to the cells below it, position the cursor over the small black box in the lower right corner of the cell. As the large white plus sign cursor () changes to a thin black one () , click and drag it downward to fill the desired cells.

Using AutoFill is also handy when you wish to copy a formula from one cell to lots of others.

Extending a Series

The Fill Handle can also be used extend a recognizable pattern such as those in a sequence of numbers, names of days, or names of months. In most cases, if the first couple of cells are filled and selected, this will be enough to establish a pattern that Excel can recognize and continue.



In this example, we selected the first two cells of an intended pattern where the number “1” was in the first cell and “2” was in the second cell. Highlighting

Numbers	Odd	Reverse	Years	Months	Days
1	1	12	2008	January	Monday
2	3	11	2009	February	Tuesday
3	5	10	2010	March	Wednesday
4	7	9	2011	April	Thursday
5	9	8	2012	May	Friday
6	11	7	2013	June	Saturday
7	13	6	2014	July	Sunday
8	15	5	2015	August	Monday
9	17	4	2016	September	Tuesday
10	19	3	2017	October	Wednesday
11	21	2	2018	November	Thursday
12	23	1	2019	December	Friday

these and dragging the fill handle downward shows the number “3” intended to be placed in the next cell to the right of the () . Dragging down farther will continue the pattern in additional cells such as in these examples:

This series fill trick can be handy if you need to sequentially number all of your observations.

Formulas and Functions

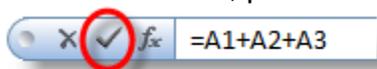
You can enter custom mathematical formulas in worksheet cells. You can also use built-in functions provided by Excel as part of these formulas. Functions and formulas can contain numbers, cell names, or cell addresses. Start a formula with an equals sign (=). By default, formulas with cell references will automatically update the calculated value if the source numbers change.

Formulas

Formulas allow you to build calculations from scratch.

1. Select the cell in which you want the result of the calculation to appear
2. Type an =
3. Type the desired formula and select one of the following to commit and run it:

Press the **Enter**, press the **Tab** key, or click the **Û** on the formula bar



As Excel worksheets are dynamic in nature, a formula automatically reevaluates the contents of the cells and displays new results once a change to a source cell is made.

Order of Operations

The order of operations is important when working with formulas in Excel. Items are treated as being read from left to right by default, and ones of higher order are processed before those of lower order.

Parentheses, Exponents, Multiplication or Division, Addition or Subtraction

Parentheses	Exponents	Multiplication	Division	Addition	Subtraction
()	^	*	/	+	-

Functions

Selecting Ranges of Cells

In working with functions, one often needs to select ranges of cells. As an example, to reference the collection of cells A1, A2, and A3, the notation **A1:A3** would be used. This has the first and last cells separated by a colon, and it indicates the use of all the cells from A1 to A3 inclusive. The cells need not be in the same row or column. For example, the range B2:D5 indicates the rectangle of cells with cell B2 in the upper left corner and cell D5 in the lower right corner.

You can directly type in the cell ranges you want to use, or you can use the mouse to select the range of cells that you want by clicking on the first cell and then dragging over the rest of the desired range.

Note that any cell in your specified range(s) that is blank will be ignored when computing values. By contrast, a cell that contains zero (0) will be included. There is,

thus, no difference between a blank cell and a zero cell in computing a sum, but a zero cell would contribute to a count function value while a blank cell would not.

Using Functions

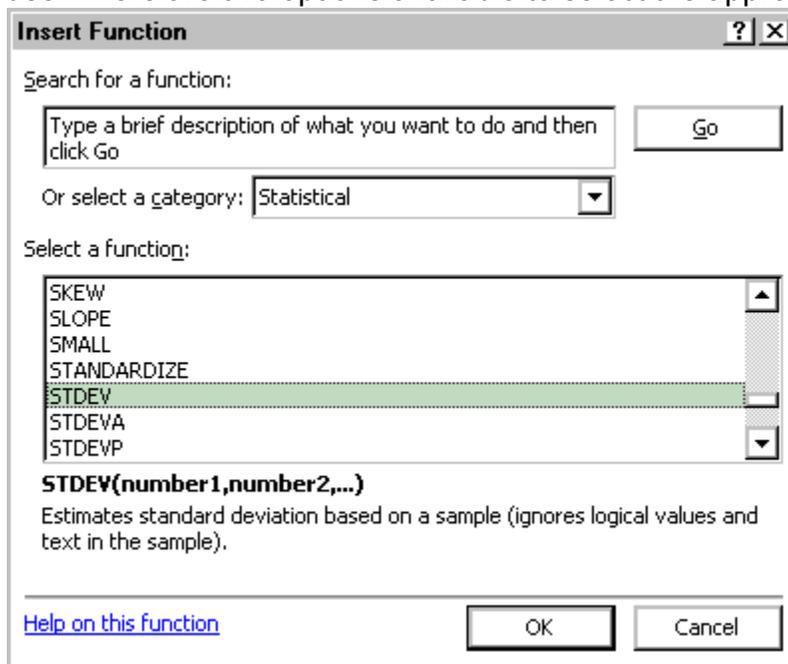
Excel has a vast array of built-in functions available that can facilitate calculations (e.g. sum, average, max, min, count). These functions may be selected from a menu, or simply typed in directly.

Using the Insert Function Menu

1. To use a function, first select the cell in which you want the answer to appear
2. Click the **Function Wizard** button

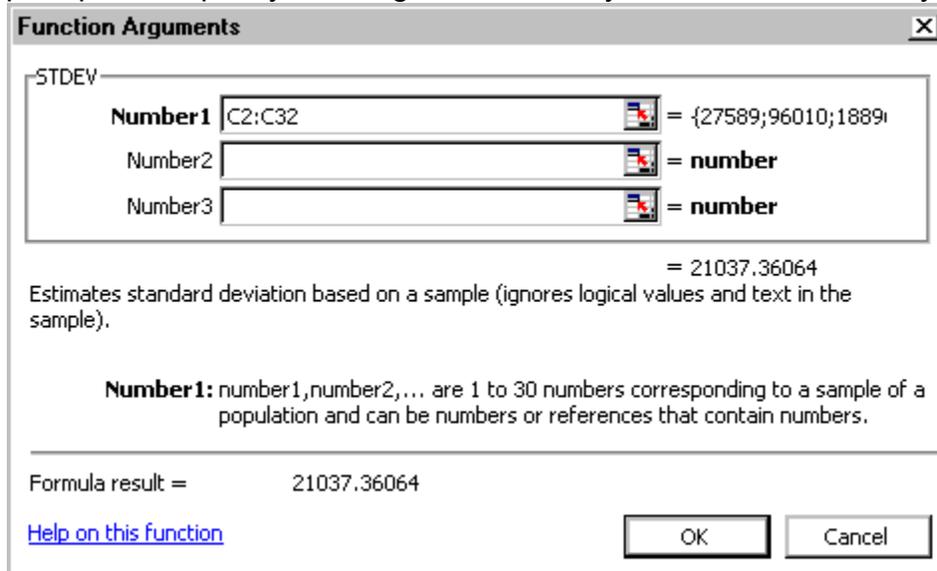


3. In the Insert Function window that appears, determine which function you wish to use. There are two options available to select the appropriate function:



- a. You can describe what you wish to do then click **Go** to let Excel suggest some possible functions.
 - b. You can select a category then choose a function from the “Select a Function” list of choices
 - c. When you click on a function name in the list, you will see more details appear.
 - d. Ask for “help on this function” if you need it
4. Click on the **OK** button

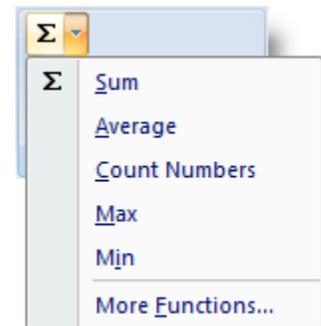
5. Follow any prompts that may appear after you select a function. You are usually prompted to specify the range of numbers you want to include in your calculation.



AutoSum

On the Home ribbon is the Σ AutoSum button. This tool will let you quickly compute a total for rows or columns of numbers.

Beside it, you will see a drop down arrow, which gives you access to some of the most commonly used functions, as well as a “More Functions” link to the Insert Functions dialog discussed above.



To use the AutoSum feature:

1. First select the cell where you want the result of the calculation to appear.
2. Click on the **AutoSum** button on the standard toolbar (the Σ symbol itself as opposed to drop down arrow).
3. Excel looks for adjacent cells which it thinks you may be trying to total up. If you require a different range, use your mouse to highlight those cells; otherwise, accept what Excel selects for you by pressing the **Enter** key. The result of the calculation will be entered into the cell you initially chose.
4. Enter the range of data or click on the red arrow button to select it from the sheet
5. Click on **OK**

Standardize Scores

=STANDARDIZE(M2,\$M\$33,\$M\$34)

	A	B	C	D	E
1	GEO_ID	CODE	P037001	P037002	P037011
2	05000US01001	01001	27,589	13,009	4,095
3	05000US01003	01003	96,010	45,983	13,020
4	05000US01005	01005	18,896	9,547	3,064
31	05000US01059	01059	20,860	9,883	2,979
32	05000US01061	01061	17,588	8,217	2,615
33		average:	27,833	13,189	4,301
34		std dev:	21037	9957	2993
35		median:	24,407	11,111	3,263

where M33 is the mean and M34 is the std dev

Practice using statistical functions

1. Enter an =
2. Enter the name of the function and (
3. Select the data or enter the range of cell addresses
4. Enter the closing) and press enter
5. Copy the formula to the right or down if you wish

7	01047	28,742	12,4
8	01049	42,740	20,2
9	01051	43,177	21,3
10	01053	25,510	12,8
11	01055	69,829	32,2
12	01057	12,579	5,8
13	01059	20,860	9,8
14	01061	17,588	8,2
	average:	=average(C2:C32)	
	std dev:		
	median:		

Examples:

=average(c2:c32)

=stdev(c2:c32)

=median(c2:c32)

About statistical analysis tools

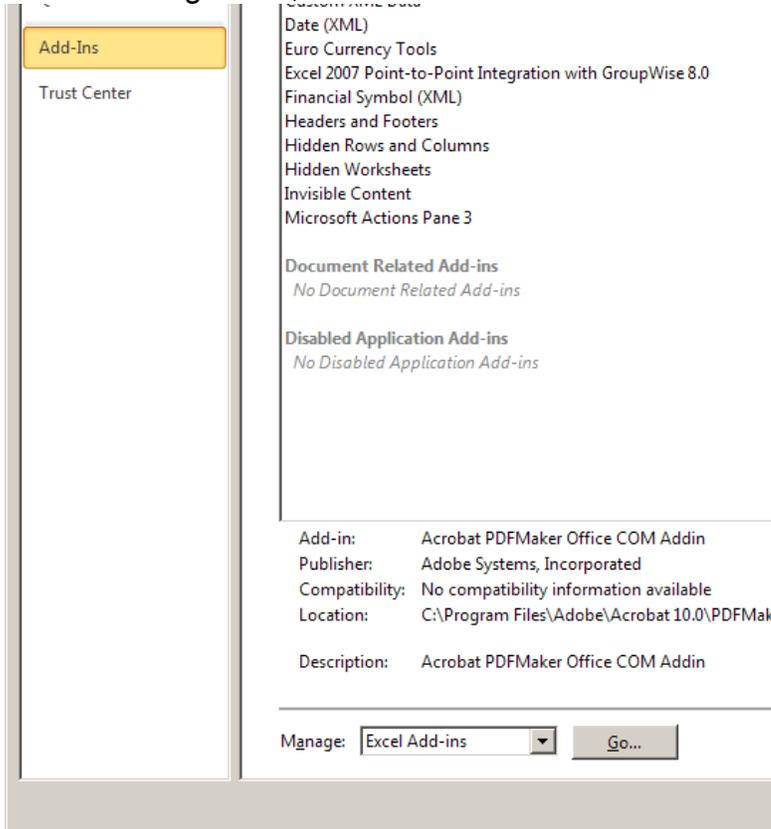
“Microsoft Excel provides a set of data analysis tools— called the Analysis ToolPak— that you can use to save steps when you develop complex statistical or engineering analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical or engineering macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Related worksheet functions Excel provides many other statistical, financial, and engineering worksheet functions. Some of the statistical functions are built-in and others become available when you install the Analysis ToolPak.”

From microsoft.com

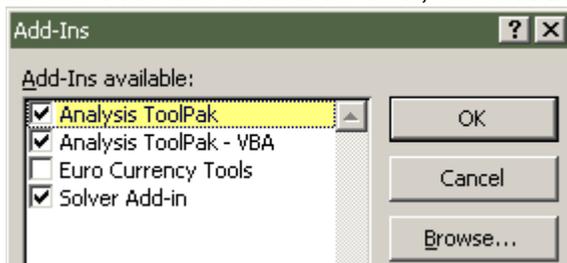
Install Analysis ToolPak

1. From the **File** tab, select **Options**.
2. Click **Add-ins**. If Analysis ToolPak is already active, it will appear in the top part of the list under "Add-ins"
3. In the "Manage" box, select **Excel Add-ins**.



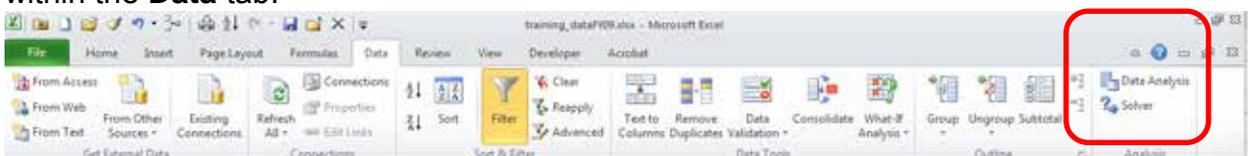
Click **Go**.

4. In the "Add-Ins available" box, check **Analysis ToolPak**, and then click **OK**.



If Analysis ToolPak is not listed in the "Add-Ins available" box, click **Browse**.

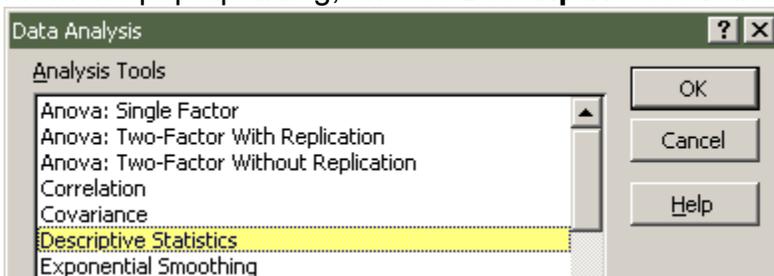
5. If you see a prompt stating that the Analysis Toolpak is not currently installed on your computer, click **Yes** to install it. This will create a "Data Analysis" section within the **Data** tab.



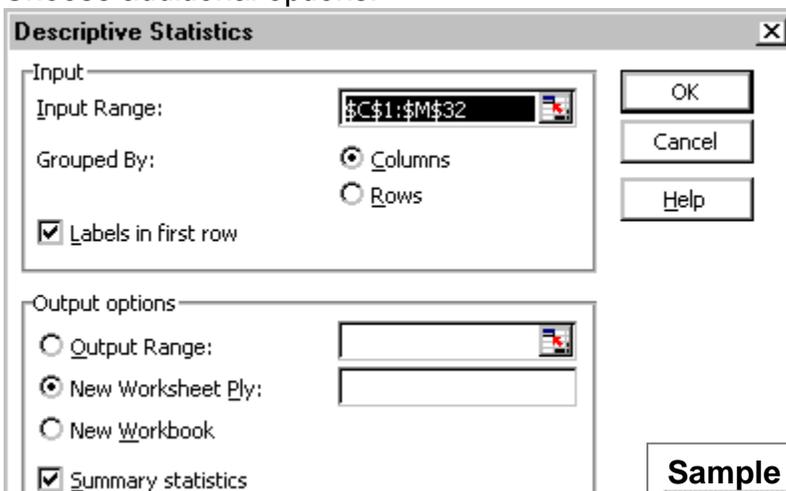
Descriptive Statistics

You can use descriptive statistics to summarize your data.

1. Click on the **Data** tab and choo
2. Choose **Data Analysis**.
3. From the pop-up dialog, choose **Descriptive Statistics** and click on **OK**



4. Enter or select the input range.
5. Choose the destination for the report (new sheet or address)
6. Choose additional options.



7. Click on **OK**.
8. You will see something like this

Sample Results	
Mean	85.67
Standard Error	1.66
Median	88
Mode	95
Standard Deviation	9.10
Sample Variance	82.85
Kurtosis	-0.34
Skewness	-0.71
Range	35
Minimum	65
Maximum	100
Sum	2570
Count	30

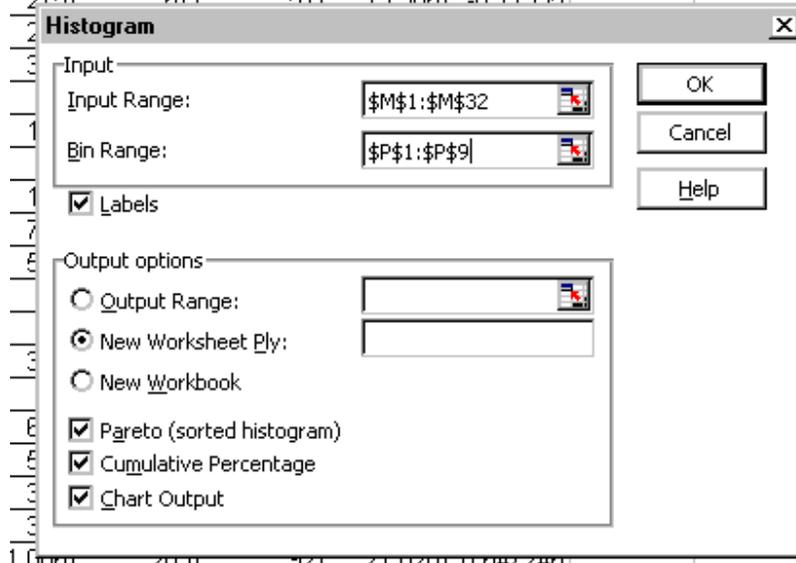
Histogram

“Histograms visually represent the frequency distribution of a data set in a graph form, allowing you to understand its statistical properties. Unlike traditional bar graphs, which often represent mean values, a histogram represents the frequency of a particular event. To make a histogram, you must have a data set that can be divided into classes, with each class having a specific frequency of responses.”

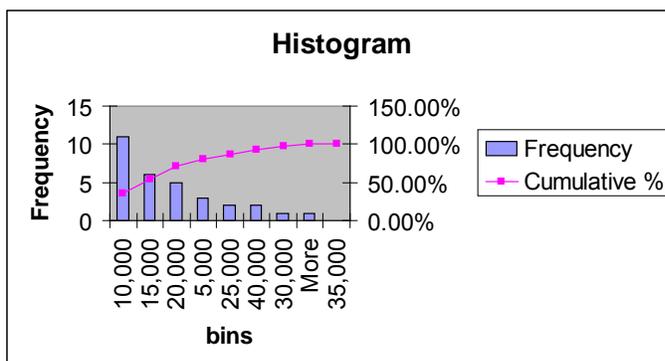
http://www.ehow.com/how_8001134_make-histogram-excel-2010.html

Data tab > Data Analysis > Histogram

J	K	L	M	N	O	P
7016	P037017	P037018	P037019	Z37019		bins
759	136	90	14,580	-0.00577		5,000
2,397	1,044	502	50,027	3.187884		10,000
130	78	12	9,349	-0.47707		15,000
94	51	15	6,599	-0.72484		20,000
338	132	44	17,216	0.23172		25,000
68	23	6	3,576	-0.9972		30,000
73	104	10	7,578	-0.63663		35,000
1,354	580	323	39,501	2.239526		40,000
275	67	51	13,386	-0.11335		



<i>bins</i>	<i>Frequency</i>	<i>Cumulative %</i>	<i>bins</i>	<i>Frequency</i>	<i>Cumulative %</i>
5,000	3	9.68%	10,000	11	35.48%
10,000	11	45.16%	15,000	6	54.84%
15,000	6	64.52%	20,000	5	70.97%
20,000	5	80.65%	5,000	3	80.65%
25,000	2	87.10%	25,000	2	87.10%



Sparklines

“Excel 2010 makes it possible to insert **sparklines**. Sparklines are graphs that fit in one cell and give you information about the data.”

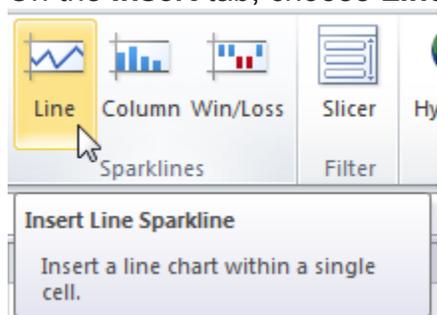
from: <http://www.easyexceltutorial.com/excel-sparklines.html>

Insert a Sparkline

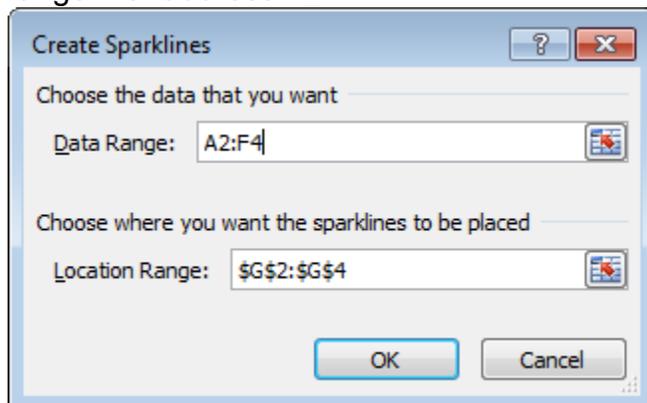
1. Select the cells where you want the sparklines to appear. In this example, we select the range with address G2:G4.

	A	B	C	D	E	F	G	
1	Data							
2	5	8	4	8	2	5		
3	4	8	7	4	9	12		
4	9	4	8	4	5	1		

2. On the **Insert** tab, choose **Line** under the Sparklines group.



3. Enter the data you want to use for the Sparklines. In this example, we use the range with address A2:F4.



Result:

	A	B	C	D	E	F	G	
1	Data							
2	5	8	4	8	2	5		
3	4	8	7	4	9	12		
4	9	4	8	4	5	1		

4. Change the value of cell C2 to 25.

Result:

	A	B	C	D	E	F	G	
1	Data							
2	5	8	25	8	2	5		
3	4	8	7	4	9	12		
4	9	4	8	4	5	1		

The sparklines change dynamically depending on the cell's value.

Customize the Sparkline

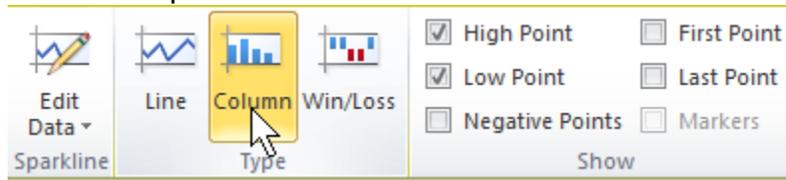
1. Select the sparklines. The Sparkline Tools contextual tab activates automatically.
2. Check High Point and Low point to highlight the highest and lowest point of data in the selected Sparkline group.



Result:

	A	B	C	D	E	F	G	
1	Data							
2	5	8	25	8	2	5		
3	4	8	7	4	9	12		
4	9	4	8	4	5	1		

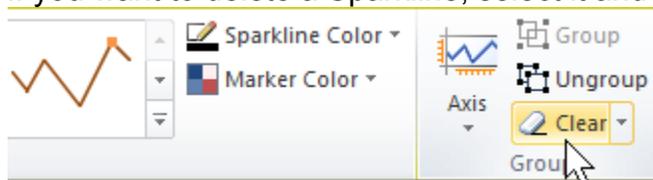
3. Click on **Column** under the Type group to change the selected Sparkline group to column sparkline.



Result:

	A	B	C	D	E	F	G	
1	Data							
2	5	8	25	8	2	5		
3	4	8	7	4	9	12		
4	9	4	8	4	5	1		

4. If you want to delete a Sparkline, select it and click on the clear button.



Chi-Square Test

Pearson's chi-squared goodness-of-fit test: CHISQ.TEST, formerly known as CHITEST

CHITEST ✖ ✔ fx =CHITEST(B2:D2,B3:D3)

	A	B	C	D	E	F	G	H	I	J
1		Red	White	Blue						
2	Observed	16	8	6						
3	expected	10	10	10						
4										
5	chi-square									
6	p>.05	=CHITEST(B3:D3)								
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										

Function Arguments

CHITEST

Actual_range B2:D2 = {16,8,6}

Expected_range B3:D3 = {10,10,10}

= 0.060810063

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Expected_range is the range of data that contains the ratio of the product of row totals and column totals to the grand total.

Formula result = 0.060810063

[Help on this function](#)

OK Cancel

In the desired cell, enter or select =CHISQ.TEST(actual_range,expected_range), where actual range holds the observed values and the expected range holds the expected values. This function will calculate the probability of getting a larger X^2 value

CHIINV ✖ ✔ fx =CHIINV(B6,2)

	A	B	C	D
1		Red	White	Blue
2	Observed	16	8	
3	expected	10	10	1
4				
5	chi-square			
6	p>.05	0.06081		
7				
8	chi-sq value	=CHIINV(E		
9				

Use the CHISQ.INV.RT function¹ to provide the actual Chi-Square value

In the desired cell, enter or select =CHISQ.INV.RT(probability,deg_freedom) where the probability is the value calculated from the Chi Square Test above and the degrees of freedom is the number of items minus 1.

Note: Use of CHISQ.TEST is most appropriate when the expected values for each cell are not too small. Some statisticians suggest that each E_{ij} should be greater than or equal to 5.

¹ Formerly CHIINV function.

T-Test

From microsoft.com: "The Two-Sample t-Test analysis tools test for equality of the population means underlying each sample. The 3 tools employ different assumptions: that the population variances are equal, that the population variances are not equal, and that the two samples represent before treatment and after treatment observations on the same subjects.

t-Test: Two-Sample Assuming Equal Variances This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with the same variances. You can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means.

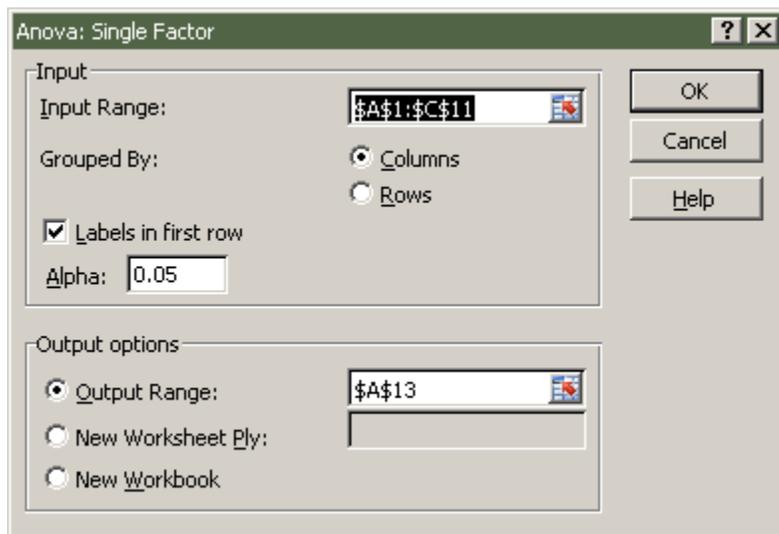
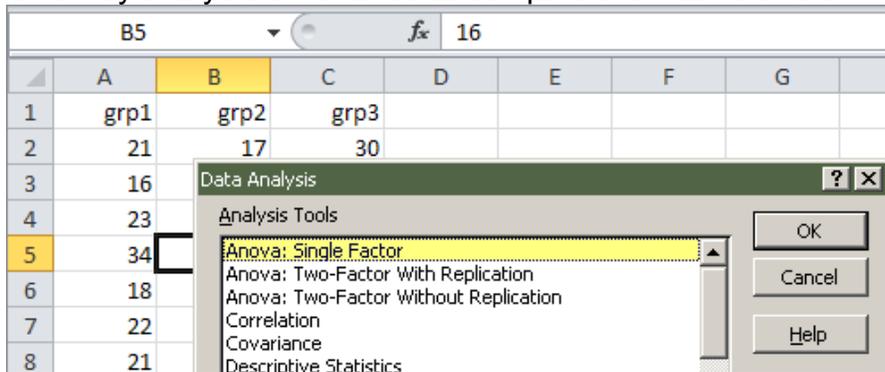
t-Test: Two-Sample Assuming Unequal Variances This analysis tool performs a two-sample student's t-test. This t-test form assumes that the two data sets came from distributions with unequal variances. You can use this t-test to determine whether the two samples are likely to have come from distributions with equal population means. Use this test when there are distinct subjects in the two samples.

t-Test: Paired Two Sample For Means You can use a paired test when there is a natural pairing of observations in the samples, such as when a sample group is tested twice: before and after an experiment. This t-test form does not assume equal variances."

t-Test: Paired Two Sample for Means		
	<i>p1</i>	<i>p2</i>
Mean	5.73	7.73
Variance	5.10	3.03
Observations	30	30
Pearson Correlation	0.76	
Hypothesized Mean Difference	0	
df	29	
t Stat	-7.49	
P(T<=t) one-tail	1.473E-08	
t Critical one-tail	1.70	
P(T<=t) two-tail	2.945E-08	
t Critical two-tail	2.05	

Anova: Single Factor

One-way analysis of variance to compare the means of 3 or more groups.



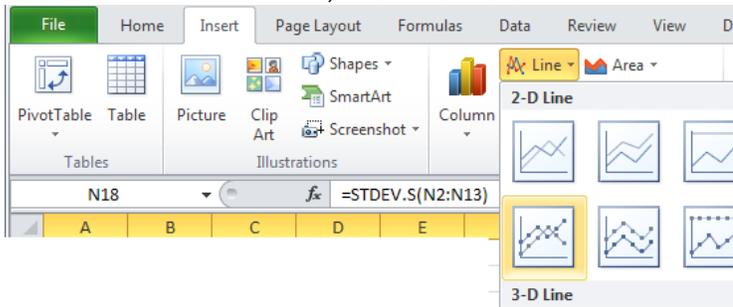
Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
grp1	10	207	20.7	28.9		
grp2	10	203	20.3	14.7		
grp3	10	301	30.1	17.2		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	615.2	2	307.6	15.18	3.82E-05	3.354
Within Groups	547.1	27	20.3			
Total	1162.3	29				

Standard Deviation Enhanced Line Graph

Calculate the mean and standard deviation for each point in time. Create an upper and lower limit line by adding and deleting one standard deviation for each mean.

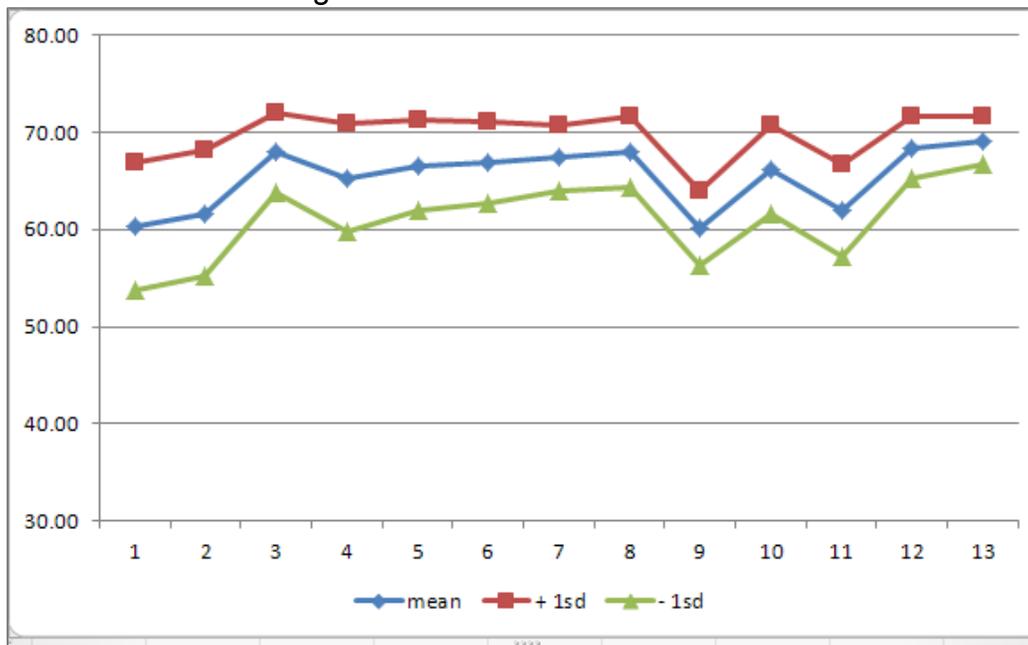
1. In cell B15, enter **=AVERAGE(B2:B13)**.
2. In cell B18, enter **=STDEV.S(B2:B13)**.
3. In cell B16, enter **=B15+B18**
4. In cell B17, enter **=B15-B18**
5. Copy and paste the formulas across the columns.
6. Select the values (do not include row 18).
7. **Insert** tab > **Line** chart, with markers

	A	B	C	D
1	id	s1	s2	s3
2	1997	59	69	72
3	8442	65	71	72
4	5114	60	70	71
5	3922	63	64	68
6	1472	46	65	71
7	8638	51	57	68
8	2615	62	60	68
9	8990	69	60	70
10	4858	57	58	69
11	2079	64	50	66
12	3873	61	62	60
13	1981	67	54	60
14				
15	mean	60.33	=AVERAGE(B2:B13)	
16	+ 1sd	66.87	=B15+B18	
17	- 1sd	53.79	=B15-B18	
18	std dev	6.54	=STDEV.S(B2:B13)	



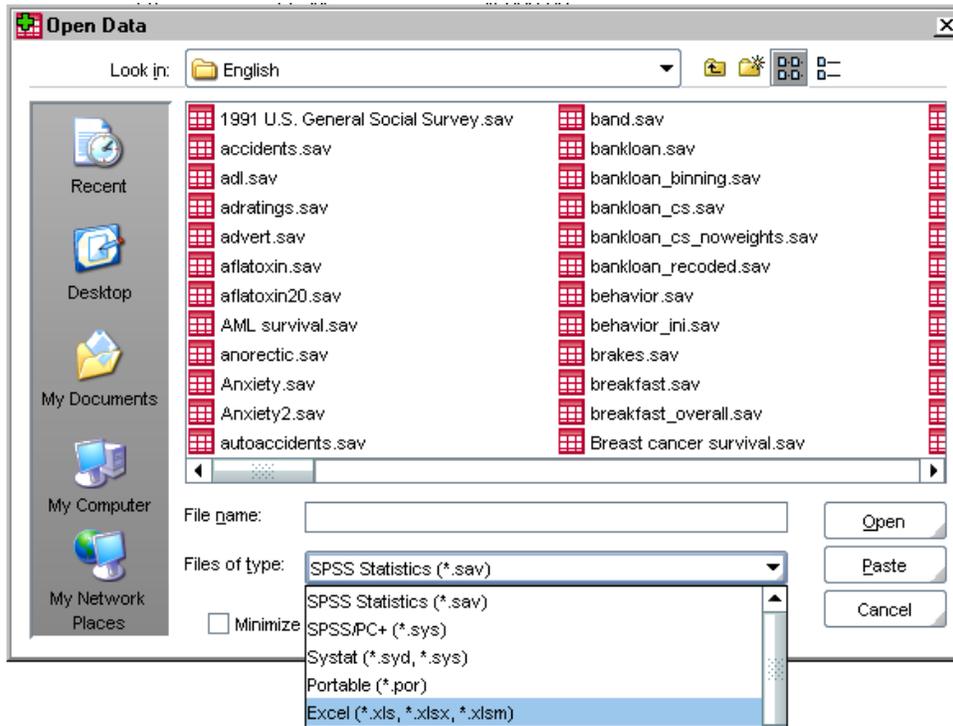
14								
15	mean	60.33	61.67	67.92	65.33	66.58	66.92	67.42
16	+ 1sd	66.87	68.15	72.04	70.83	71.28	71.13	70.81
17	- 1sd	53.79	55.18	63.79	59.83	61.88	62.71	64.02

8. You will see something like

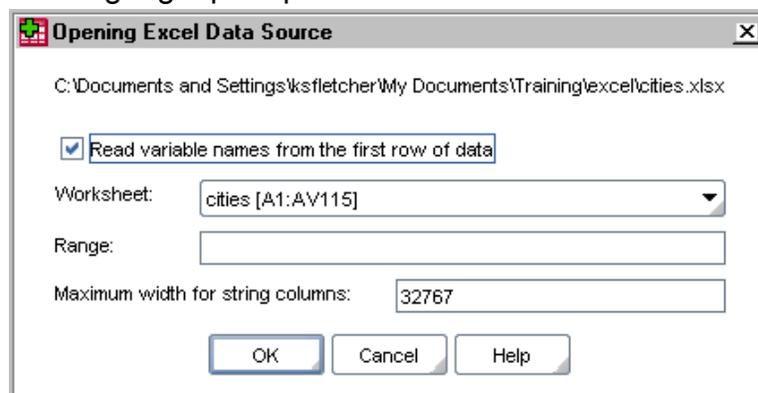


Import Excel Data into SPSS

1. Go to the **File** menu and select **Open > Data**.
2. Change the location in the "Look in" box to the subdirectory where your file is.
3. Change the "Files of type" selection to look for Excel (*.xls) files.



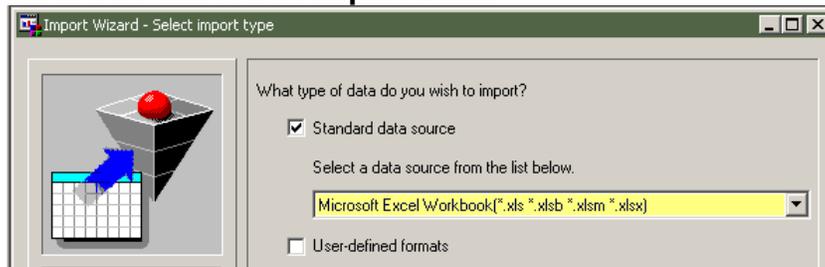
4. Select the file.
5. You might get prompted about the variable names:



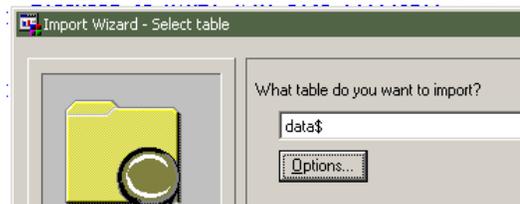
6. Click on **OK**. You will see the data appear in the Data Editor window.
7. You may need to modify some of the variable definitions (in Variable View)

Import Excel Data into SAS

1. Start in your Editor window.
2. Go to **File** and select **Import Data**.



3. Select **Excel Workbook** from the dropdown list. Click on **Next**.
4. Click on the **Browse** button and select the file.
You might need to tell it to look for .xlsx files.
5. Select the table if there is more than one sheet in the workbook:



6. Click on the **Options** button if you need to supply additional info.
7. Click on **Next**.
8. Give the library member a desired dataset name.
9. Click on **Finish**.(skip the step that offers to save the commands to import).

References

- **Data Analysis with Spreadsheets**, Patterson & Basham, Pearson, 2006
- **About Statistical Analysis Tools**, Excel 2003: <http://office.microsoft.com/en-us/excel-help/about-statistical-analysis-tools-HP005203873.aspx>
- **Microsoft Excel Blog**: <http://blogs.office.com/b/microsoft-excel/>
 - <http://blogs.office.com/b/microsoft-excel/archive/2009/09/10/function-improvements-in-excel-2010.aspx>
 - <http://blogs.office.com/b/microsoft-excel/archive/2009/09/14/function-consistency-improvements-in-excel-2010.aspx>
- **Excel Skills Builder Video Training**
<http://office.microsoft.com/en-us/excel/excel-skills-builder-FX102592909.aspx>
- **Excel Tutorial – Data Analysis**
<http://www.easyexceltutorial.com/excel-data-analysis.html>
- **How to Make a Histogram in Excel 2010**:
http://www.ehow.com/how_8001134_make-histogram-excel-2010.html
- **Michael Girvin, Business Courses & Excel is Fun**
<http://people.highline.edu/mgirvin/>