
A Quick Guide to Using Excel 2007's Regression Analysis Tool



Table of Contents

	Page
The Analysis Toolpak	1
If the Add-In is ALREADY Installed	1
If the Add-In is NOT ALREADY Installed	1
Run a Multiple Regression: An Example	
Introduction.....	2
Steps to Follow	2
1. Rearrange the Data as Necessary	2
2. Open the Regression Analysis Tool	3
3. Complete the "Regression" Dialog Box.....	4
4. Run the Regression Analysis and View the Results	5
The Regression Results Areas.....	6
Run Another Regression Analysis	6

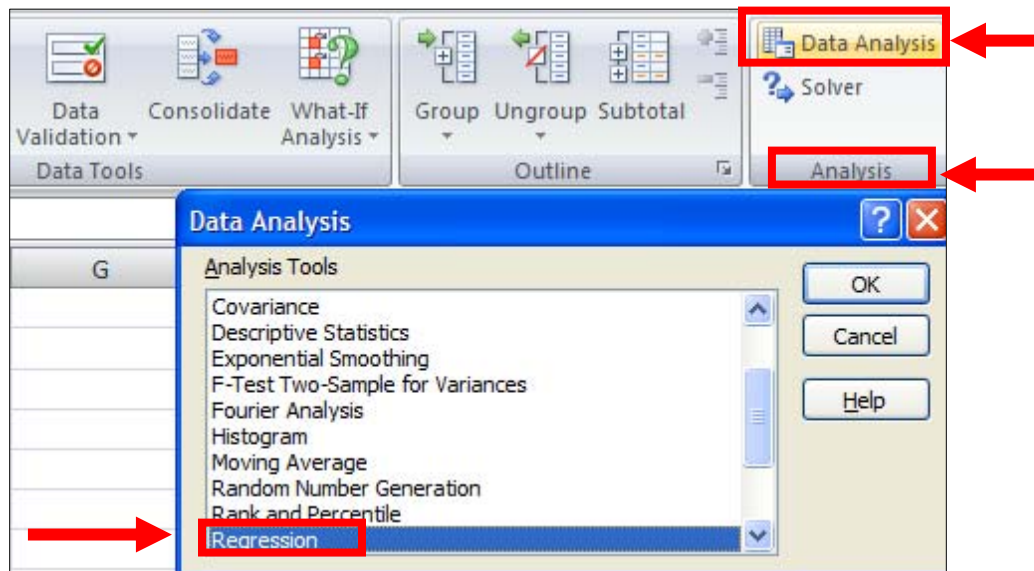
Using Excel 2007's Regression Analysis Tool

The Analysis Toolpak

Excel 2007 has a built-in regression analysis tool that's packaged as part of its "Analysis Toolpak". The Analysis Toolpak is a standard component of Excel. Microsoft makes it available as an Excel add-in. If you loaded your copy of Excel on your computer, you may recall that you had a choice during the installation process whether or not to include add-ins such as the Analysis Toolpak.

If the Add-In is ALREADY Installed

If the Analysis Toolpak add-in is already installed, you'll see an "Analysis" group in the "Data" tab of Excel's ribbon. Click the "Data Analysis" button in that group to open the "Data Analysis" dialog. Then scroll down to find the "Regression" option.



The Analysis Toolpak add-in is installed in Excel 2007 on all FuquaNet MBA PCs.

If the Add-In is NOT ALREADY Installed

If you're using your own computer and you *do not yet have the add-in installed*, follow these steps to install it:

1. Close Excel if you have it open.
2. Make your installation medium (probably a CD) available to your computer.
3. Restart the Excel installation routine. (Note that you *need not* reinstall Excel entirely.)
4. Find the option to modify the installation and choose the Analysis Toolpak as an add-in that should be activated.
5. Exit the installation process.
6. Open Excel and confirm that the "Data Analysis" dialog is available (see above).

Run a Multiple Regression: An Example

Introduction

This example walks you through how to use Excel 2007's built-in regression tool to analyze whether information collected by the Mrs. Smyth's Gourmet Frozen Fruit Pie Company (price, advertising, competitors' pricing, etc.) can predict company pie sales. The worksheet with pie data is partially shown in the illustration below. There are ten columns of data with forty-eight records.

City	Yr/Qtr	Row	Unit Sales	(\$) Price	(\$) Ads	(\$) Comp_Price	(\$) Income	Population	(T) Time Var
Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1
	2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2
	2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3
	2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4
	2004-1	5	282,118	6.36	18,815	6.36	33,879	4,278,912	5
	2004-2	6	203,396	5.98	14,176	4.88	34,186	4,359,442	6
	2004-3	7	167,447	6.64	17,030	5.22	35,691	4,363,494	7
	2004-4	8	361,677	5.30	14,456	5.80	35,950	4,380,084	8
Chicago, IL -	2003-1	9	401,805	6.08	27,183	4.99	34,983	9,184,926	1
Gary, IN -	2003-2	10	412,312	6.13	27,572	6.13	35,804	9,237,683	2
Kenosha, WI	2003-3	11	321,972	7.24	34,367	5.82	35,898	9,254,182	3
	2003-4	12	445,236	6.08	26,895	6.05	36,113	9,272,758	4
	2004-1	13	479,713	6.40	30,539	5.37	36,252	9,300,401	5
	2004-2	14	459,379	6.00	26,679	4.86	36,449	9,322,168	6
	2004-3	15	444,040	5.96	26,607	5.29	37,327	9,323,331	7
	2004-4	16	376,046	7.21	32,760	4.89	37,841	9,348,725	8
Dallas-Fort Worth, TX	2003-1	17	255,203	6.55	19,880	6.97	34,870	5,294,645	1
	2003-2	18	270,881	6.11	19,151	6.25	35,464	5,335,816	2
	2003-3	19	330,271	5.62	15,743	6.03	35,972	5,386,134	3

For this example regression operation:

We use as the predictor variables (the independent or X variables) the data for **Price, Ads, Comp_Price, Income, Population, and Time Var.**

We use as the predicted variable (the dependent or Y variable) the data for **Unit Sales.** Unit Sales holds actual sales data.

Steps to Follow When Running a Regression with Excel

1. Rearrange the Data as Necessary

With Excel's regression tool the independent X variables you use in your analysis must be *located together* in the worksheet. There must be no blank columns or columns with non-relevant data interrupting the range of X variables. The dependent Y variable need not be located adjacent to the X variables, but all Y variable values must be in a single range. These location specifications are required by Excel's built-in regression tool.

For the data in our sample, the data need not be rearranged. The X variables are together in Columns F through K and the Y variable values are all in Column E.

	B	C	D	E	F	G	H	I	J	K
Mrs. Smyth's Gourmet Frozen Fruit Pie Regional Market Demand Data, 2003-1 to 2										
					(\$)	(\$)	(\$)	(\$)		(T)
	City	Yr/Qtr	Row	Unit Sales	Price	Ads	Comp Price	Income	Population	Time Var
	Atlanta, GA	2003-1	1	193,334	6.39	15,827	6.92	33,337	4,116,250	1
		2003-2	2	170,041	7.21	20,819	4.84	33,390	4,140,338	2
		2003-3	3	247,709	5.75	14,062	5.28	33,599	4,218,965	3
		2003-4	4	183,259	6.75	16,973	6.17	33,797	4,226,070	4
		2004-1	5	282,118	6.36	18,815	6.36	33,879	4,278,912	5

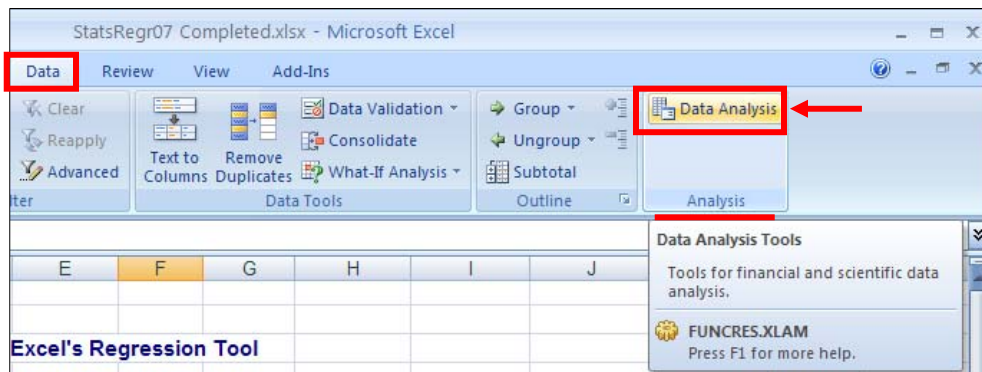
Columns not used in this regression.

The dependent, Y variable.

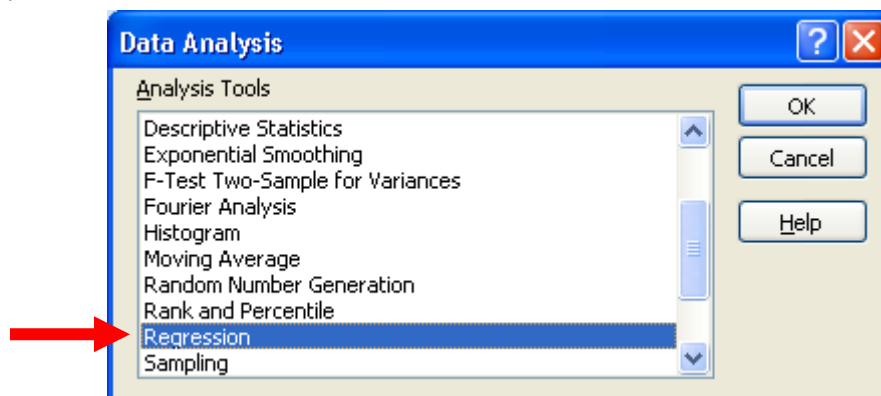
Independent, X variables, located together in the worksheet.

2. Open the Regression Analysis Tool

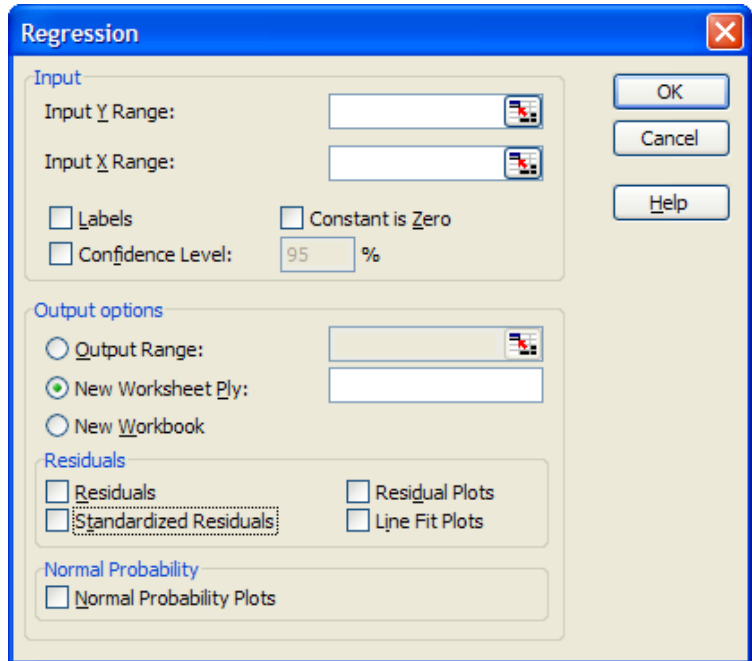
In the "Data" tab on Excel's ribbon find the "Analysis" group and click the "Data Analysis" button.



Excel opens its "Data Analysis" dialog. Scroll down and find the "Regression" option. Click OK.



Excel 2007's "Regression" dialog displays. This is the sole interface to Excel's regression tool. All your interaction with Excel's regression tool happens in this dialog.



3. Complete the "Regression" Dialog Box

	E	F	G	H	I	J	K
	rozen Fruit Pie Regional Market Demand Data, 2003-1 to 2						
		(\$)	(\$)	(\$)	(\$)		(\$)
	Unit Sales	Price	Ads	Comp Price	Income	Population	Time Var
	193,334	6.39	15,827	6.92	33,337	4,116,250	1
	170,041	7.21	20,819	4.84	33,390	4,140,338	2
	247,709	5.75	14,062	5.28	33,599	4,218,965	3

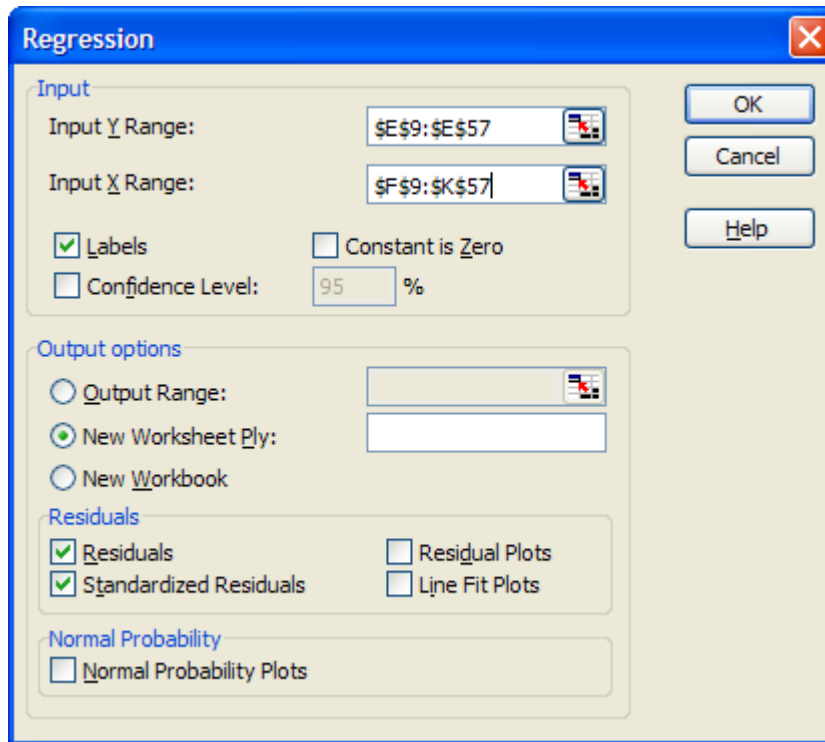
- Specify "Unit Sales" as the "Input Y Range". Include the column header cell (the cell that holds the text "Unit Sales") in your range specification¹.
- Specify the range of data in Columns F through K as the "Input X Range". Include only *one* row of column headers with your range specification (that is, include only the labels directly above the data).
- In the "Regression" dialog box make sure the *Labels* box is checked. With this option checked Excel will recognize that the first cell in each column of data is a data label.

¹ In this selection operation be careful to select *data and header only* and not an entire column. Here's an easy way to select just the contiguous data from a column of data: Click the top cell. Depress the SHIFT key. Tap the End key and then tap the down arrow key. Excel selects all data down the column until it reaches the first empty cell in the column.

- Direct Excel to put its results on a new worksheet by choosing the option *New Worksheet Ply*.

- For outputs, select *Residuals*, and *Standardized Residuals*.

With data ranges and choices specified, your dialog should now look like the one below.



4. Run the Regression Analysis and View the Results

With the dialog box options completed as described above, click *OK* to execute the regression.

The *New Worksheet Ply* default output option means that Excel locates the regression outputs on a new, separate worksheet that it creates and puts to the left of the worksheet that holds the original data.

Excel completes the regression analysis leaving a large range of the worksheet selected. Click any cell to turn off the range selection. Because the Regression tool generates many outputs on a single sheet, you may want to use Excel's Zoom Control option to reduce the new worksheet size in order to get a view of how the outputs are arranged.

	A	B	C	D	E	F	G	H	I	J
1	SUMMARY OUTPUT									
2										
3	Regression Statistics									
4	Multiple R	0.94656								
5	R Square	0.895975								
6	Adjusted R Square	0.880752								
7	Standard Error	60699.81								
8	Observations	48								
9										
10	ANOVA									
11		df	SS	MS	F	Significance F				
12	Regression	6	1.3E+12	2.17E+11	58.85609	1.39E-18				
13	Residual	41	1.51E+11	3.68E+09						
14	Total	47	1.45E+12							
15										
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
17	Intercept	646957.8	154146.9	4.197022	0.000142	336651.9	958263.8	336651.9	958263.8	
18	Price	-127443	15112.6	-8.43288	1.7E-10	-157963	-96922.3	-157963	-96922.3	
19	Ads	5.352977	1.113918	4.805641	2.09E-05	3.103374	7.602579	3.103374	7.602579	
20	Comp_Pric	29337.34	12388.59	2.368093	0.022677	4318.069	54356.61	4318.069	54356.61	
21	Income	0.344133	3.186721	0.10799	0.914531	-6.09158	6.779846	-6.09158	6.779846	
22	Population	0.023966	0.002349	10.20232	8.12E-13	0.019222	0.02871	0.019222	0.02871	
23	Time Var	4405.94	4400.397	1.00126	0.322574	-4480.84	13292.72	-4480.84	13292.72	
24										
25										
26										
27	RESIDUAL OUTPUT									
28										

Widen the columns on the new worksheet in order to see all the text results properly. A quick way to do this is to select the blank square at the intersection of the row numbers and column headers (to select the entire worksheet), and then double-click the dividing line between any two column headers. Excel widens each column so the widest entry in that column displays completely.

The Regression Results Areas

Selecting the *Residuals*, and *Standardized Residuals* options in the “Regression” dialog produces the output partially shown below.

In this view columns have been widened to show the results completely.

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0.946559629								
R Square	0.895975132								
Adjusted R Square	0.880751981								
Standard Error	60699.8107								
Observations	48								
ANOVA									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	6	1.30112E+12	2.16853E+11	58.8560875	1.39375E-18				
Residual	41	1.51063E+11	3684467019						
Total	47	1.45218E+12							
<i>Coefficients</i>									
Intercept	646957.8174	154146.8797	4.197021819	0.00014152	335651.8818	958263.753	335651.8818	958263.753	
Price	-127442.8033	15112.59845	-8.432884903	1.699E-10	-157963.3147	-96922.2919	-157963.315	-96922.2919	
Ads	5.352976863	1.113917711	4.805540671	2.0919E-05	3.103374433	7.602579293	3.103374433	7.602579293	
Comp_Price	29337.34026	12388.5933	2.368092934	0.02267716	4318.0688	54356.61171	4318.0688	54356.61171	
Income	0.344132846	3.186720717	0.107989647	0.91453065	-6.091580132	6.779845825	-6.09158013	6.779845825	
Population	0.023965666	0.002349041	10.20232045	8.1243E-13	0.019221682	0.02870965	0.019221682	0.02870965	
Time Var	4405.940135	4400.39741	1.001259596	0.32257427	-4480.842623	13292.72289	-4480.84262	13292.72289	
RESIDUAL OUTPUT									
	<i>Observation</i>	<i>Predicted Unit Sales</i>	<i>Residuals</i>	<i>Standard Residuals</i>					
	1	234861.2336	-41527.23358	-0.732491627					
	2	101059.9918	68981.00824	1.216743968					
	3	270227.062	-22518.06199	-0.397192166					
	4	189121.3616	-5862.361642	-0.103405174					
	5	259958.8857	22159.11427	0.39006075					

Excel labels the output areas “SUMMARY OUTPUT”, “ANOVA”, and “RESIDUAL OUTPUT”.

Run Another Regression Analysis

Perhaps after viewing the results of the first analysis you decide you’d like to run another regression analysis using different variables or a different combination of variables. To do so, go through the same steps as above. Excel adds a new worksheet to hold the new results. The original results (from the first analysis) remain in place.

End