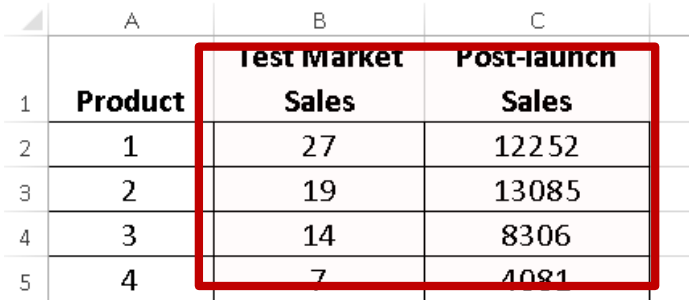
Generate a Regression Equation

Excel Step-by-Step How-to for Windows and Excel for Mac 2016 (v.16) or later

Excel for Mac 2011 (v.14) Instructions on page 12

Instructions: Use this guide to generate a regression equation using Excel. Three methods are described: creating a scatter plot, using the built-in Excel functions, and running a regression report using statistics tools.

Data requirement: Two variables (one independent, one dependent), quantitative data
Sample data: Test sales and post-launch sales

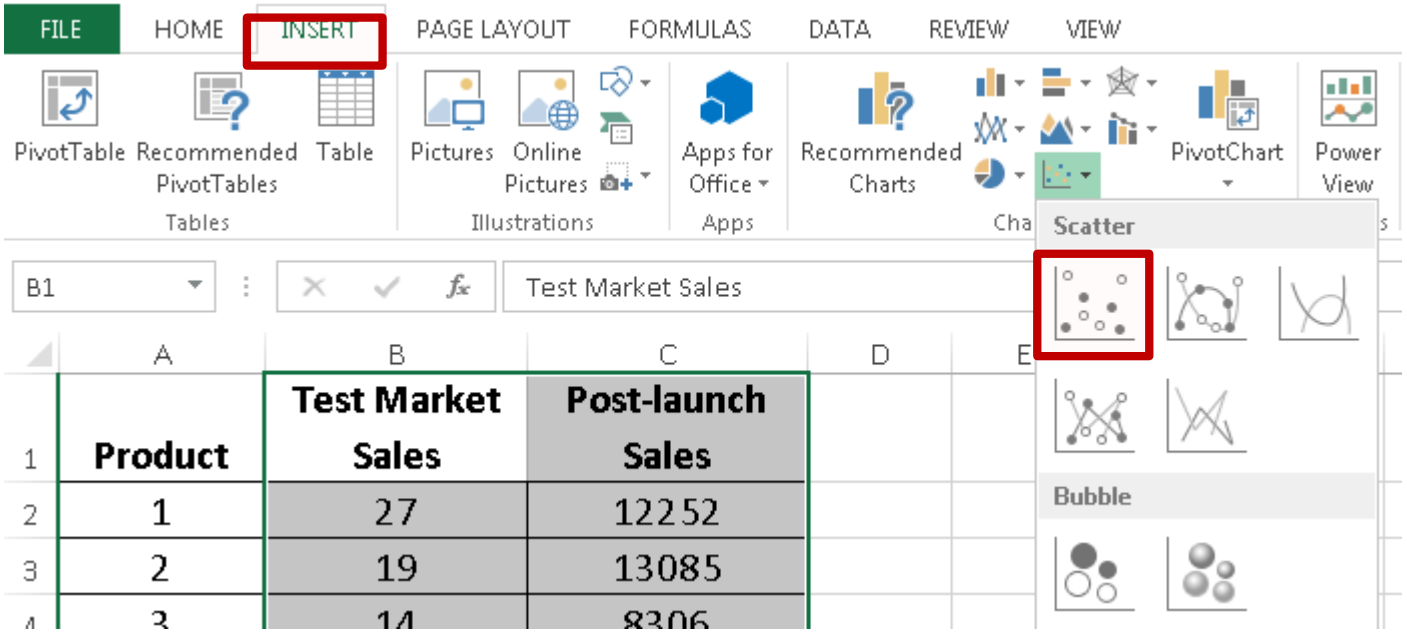
Step	Windows Instructions + Screen Shot
	Scattler Plot Method
1. Arrange the data into two adjacent columns so that the independent variable is in the left column and the dependent variable is in the right column.	

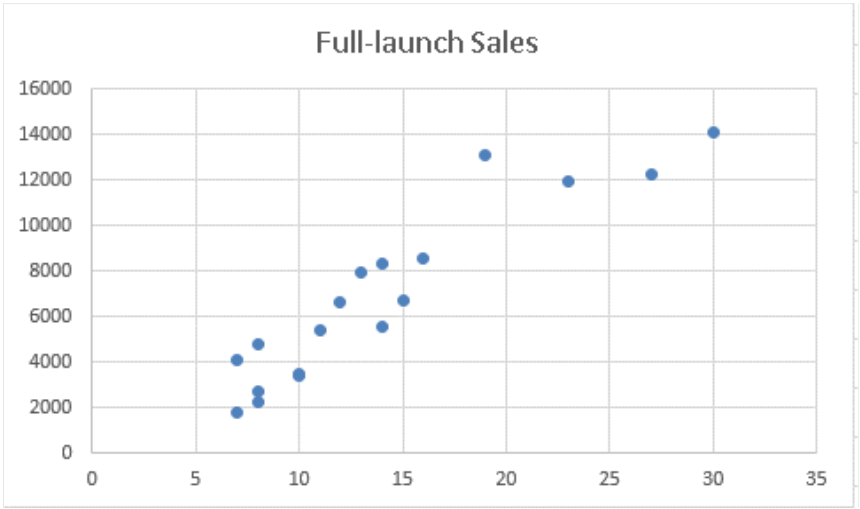
	A	B	C
1	Product	Test Market Sales	Post-launch Sales
2	1	27	12252
3	2	19	13085
4	3	14	8306
5	4	7	4081

In this example, the data are in columns B and C. The values in column B correspond to the X-axis (independent variable) and the column C data correspond to the Y-axis (dependent variable).
Note: Data for your dependent variable should always be the in right column.

2. Select your data and insert a Scatter (X,Y) chart.

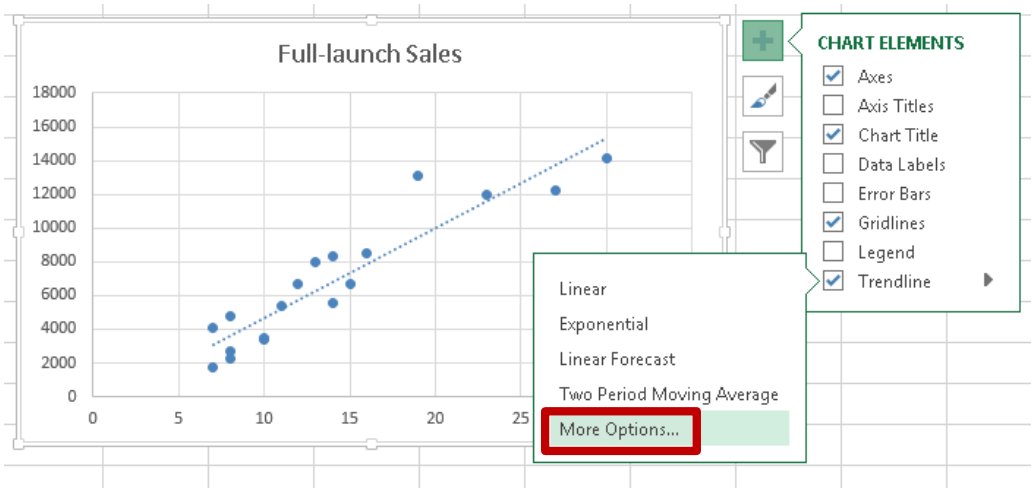
After selecting the data in both columns, including the header rows, select the Insert tab. Navigate to the Scatter (X,Y) chart icon and select it. Excel always uses the first column for the x-axis and the second column for the y-axis. If your dependent variable data is in the right column, it will always correspond to the y-axis.





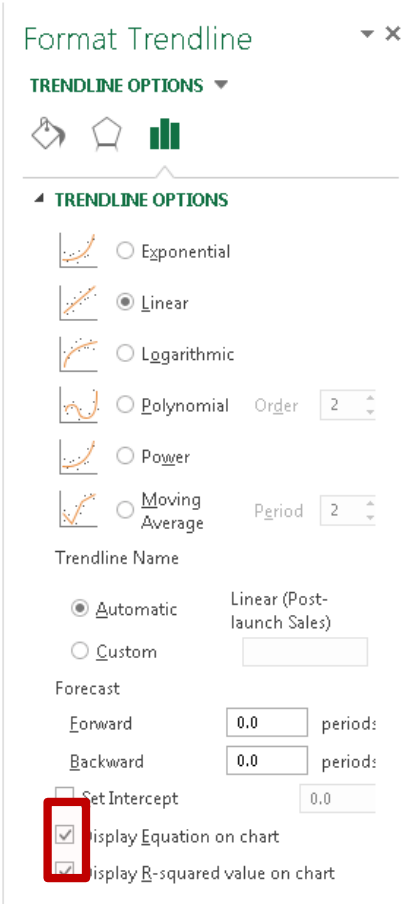
3. Display the best fit line on the scatter plot.

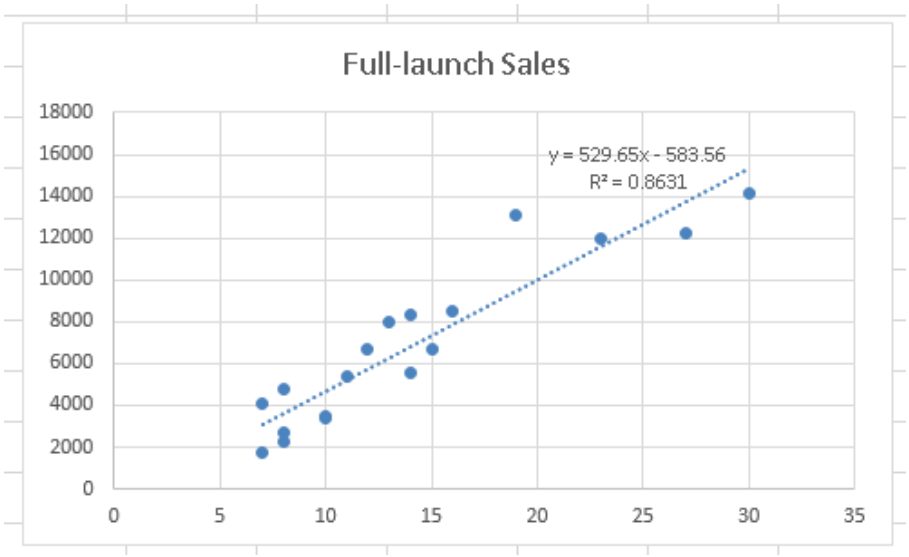
Select (right click) any point in the scatter plot to bring up the Chart Elements menu. At the bottom of this menu, select Trendline. After adding the trendline, select More Options.



4. Display the equation and R-squared value of the trendline on the chart.

Navigate to the Format Trendline pane, and select Display Equation on chart and Display R-squared value on chart.





Excel Function Method

1. Label area to calculate elements of regression equation.

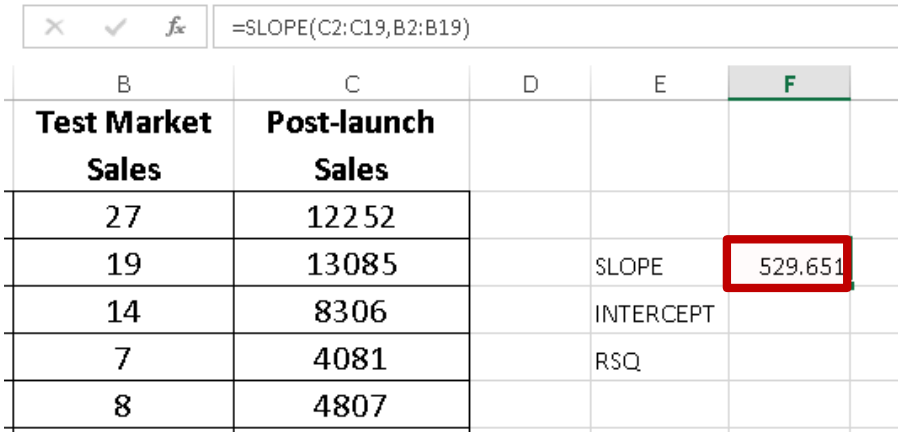
Label cells for the slope, intercept, and the R-squared value in the sheet.

Test Market Sales	Post-launch Sales			
27	12252			
19	13085		SLOPE	
14	8306		INTERCEPT	
7	4081		RSQ	
8	4807			

2. Use the SLOPE function to calculate the slope.

=SLOPE(*known y's*, *known x's*)

Next to the cell labeled SLOPE, enter the SLOPE function into the function box. Select the column of dependent data for the first parameter of the function and select the column of independent data for the second parameter. For this example, the known y's are populated in the cell range C2:C19, so that is the range for that parameter of the formula.



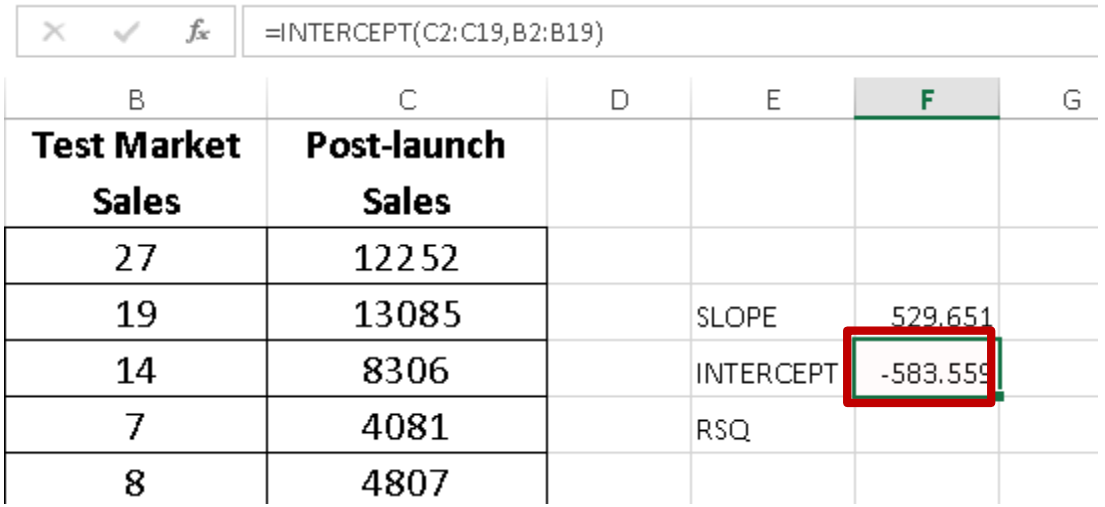
TIP: To select a column of data

TIP: When identifying a data range, to select a block of populated cells in a column of data you can first select the top row in that block. The cell will be surrounded by a highlight. Then double-click in the solid square in the lower right corner of that highlight.

Another method to efficiently collect a column of data is to highlight the top cell of the column that you want and press CTRL + Shift + ↓. This will select everything between your highlighted cell and that first blank row.

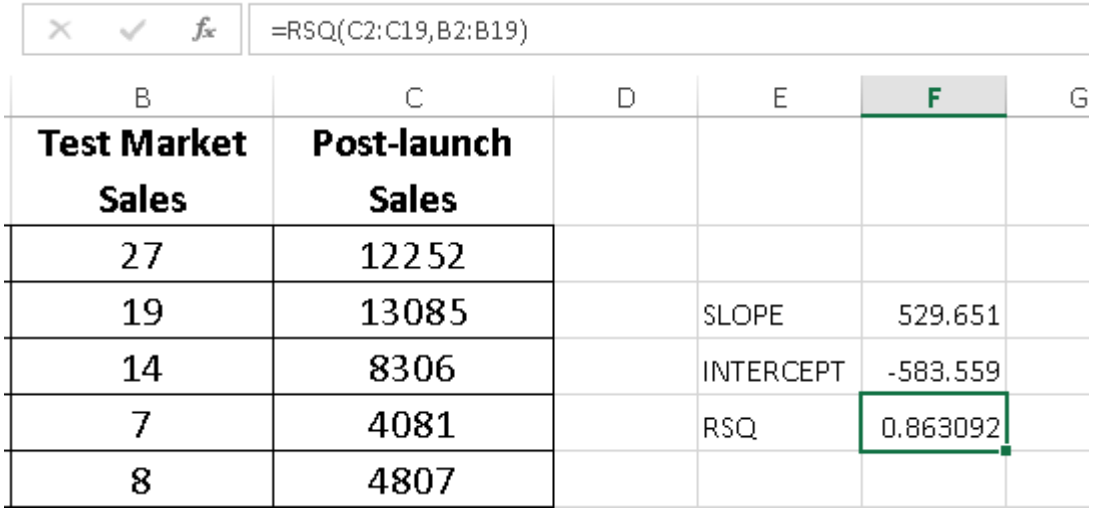
3. Use the INTERCEPT formula to calculate the intercept.

=INTERCEPT(*known y's*, *known x's*)
Type the INTERCEPT formula into the formula box. Select the column of dependent data for the first parameter of the formula and select the column of independent data for the second parameter the same way as for the SLOPE formula.



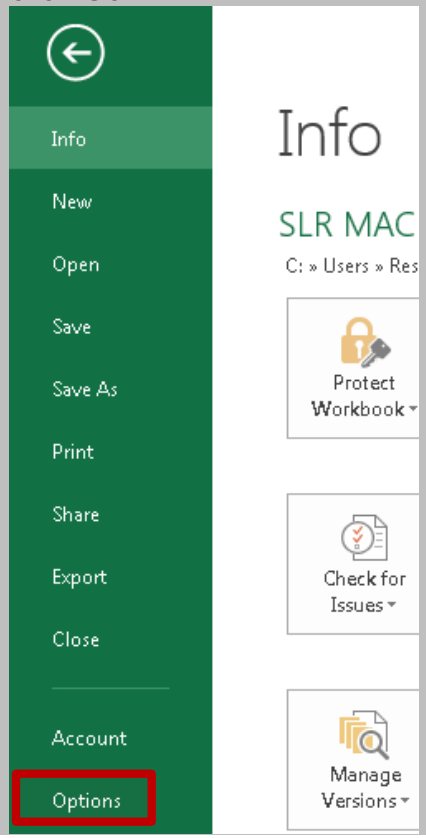
4. Use the RSQ formula to calculate the R-squared value.

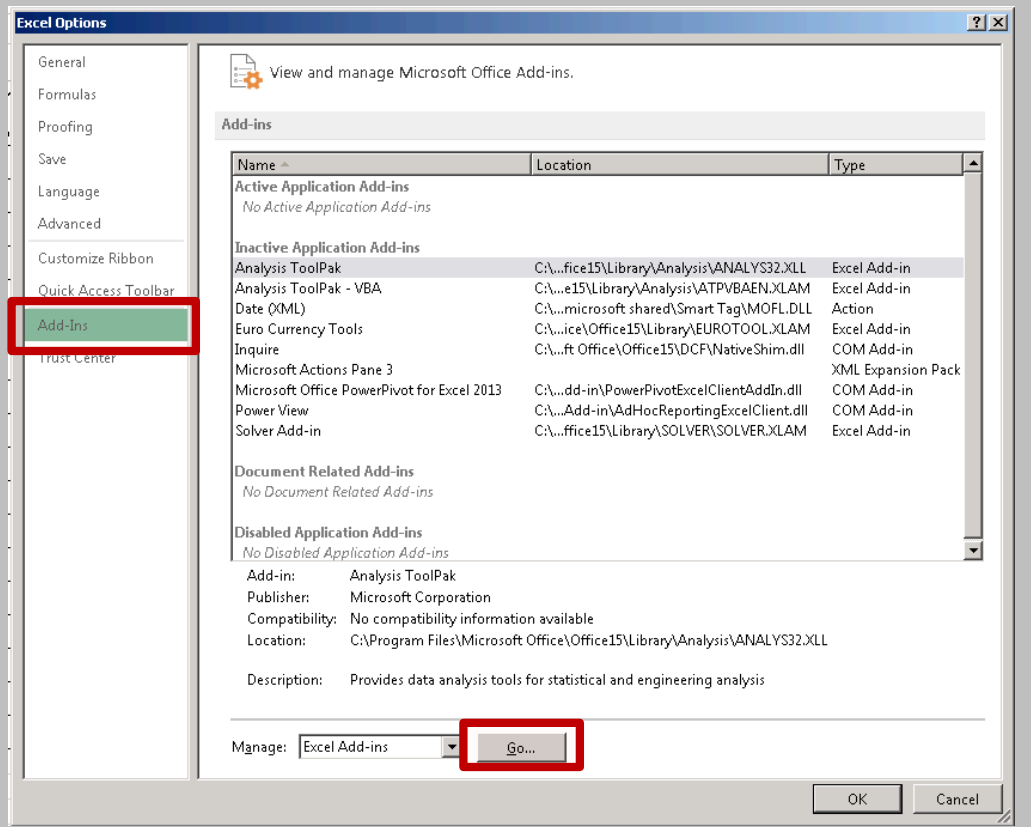
=RSQ(*known y's*, *known x's*)
Type the RSQ formula into the formula box. Select the column of dependent data for the first parameter of the formula and select the column of independent data for the second parameter the same way as for the SLOPE and INTERCEPT formula.



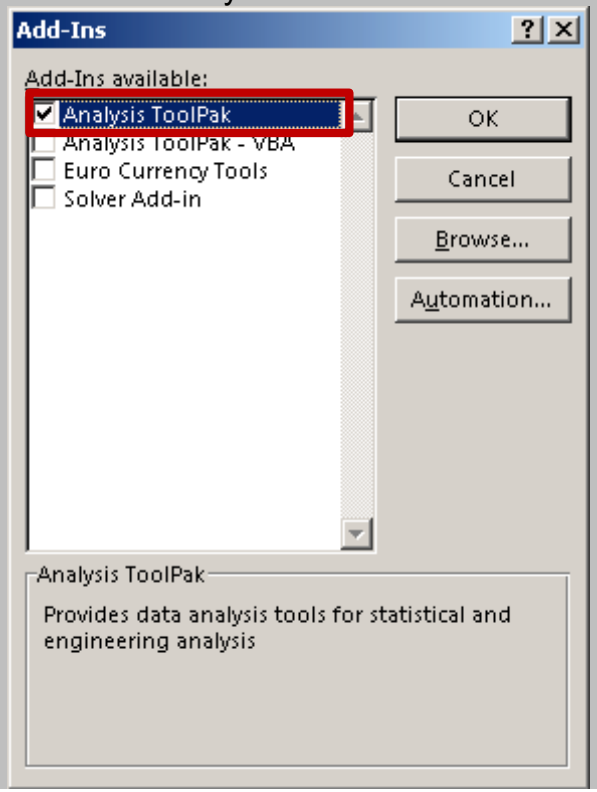
Data Analysis Method

1. Set up a third-party data analysis tool	Navigate to File and select Options. In the Excel Options pop-up window, navigate to Add-Ins and click Go.



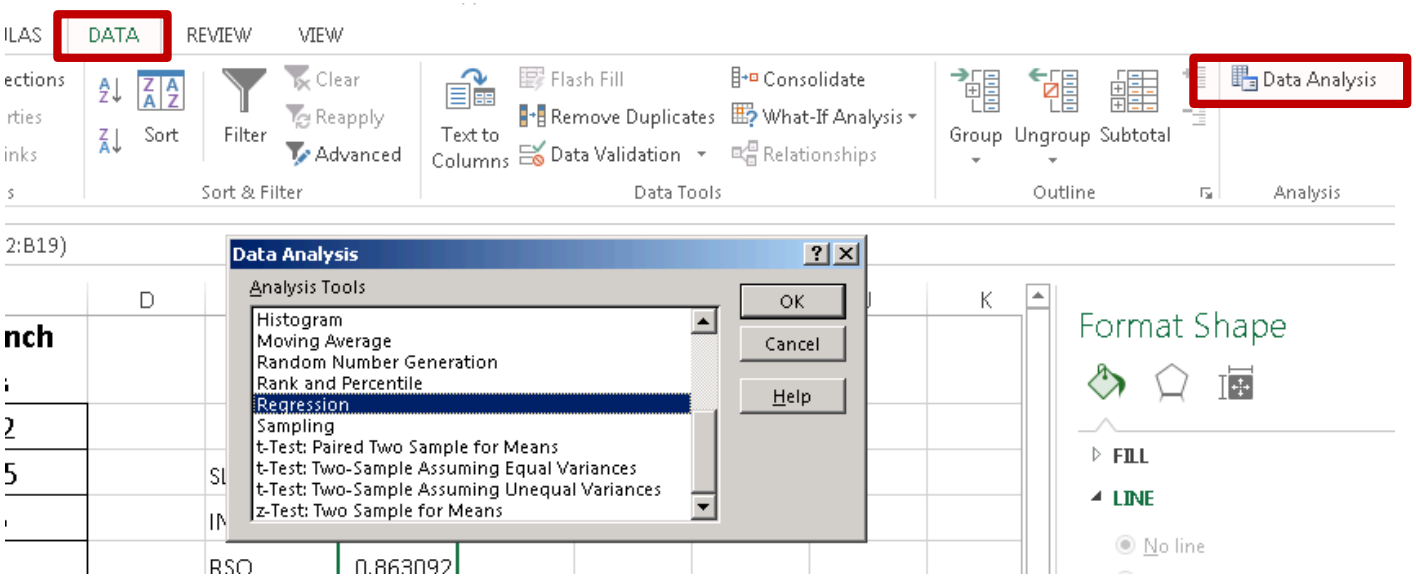


Select the Analysis TookPak and click OK.



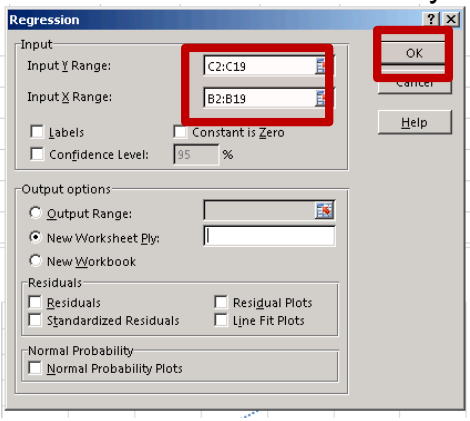
1. Open the Data Analysis tool.

Navigate to the Data ribbon and select the Data Analysis icon on the right. Select Regression and click OK.



2. Enter data.

Enter the range that holds the dependent data in the Y field and enter the range that hold the independent data in the X field. Click OK. By default, Excel creates a new worksheet when a regression is run.



3. View the regression table.

The R-squared value is found under Regression Statistics. The intercept and slope can be found in the bottom table.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.929028
R Square	0.863092
Adjusted R Square	0.854536
Standard Error	1480.396
Observations	18

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	1	2.21E+08	2.21E+08	100.8672	2.59E-08
Residual	16	35065176	2191574		
Total	17	2.56E+08			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-583.559	816.6188	-0.7146	0.485151	-2314.71	1147.596	-2314.71	1147.596
X Variable 1	529.651	52.73694	10.04326	2.59E-08	417.8537	641.4483	417.8537	641.4483

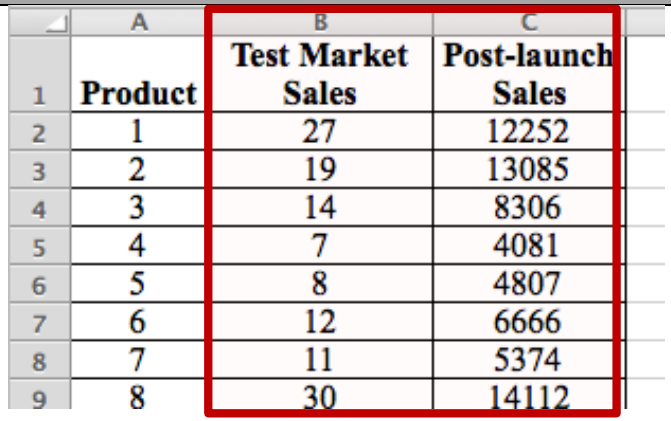
Create a Regression Equation

Excel Step-by-Step How-to for Mac Excel 2011 (v.14) or earlier

Excel for Windows and Excel for Mac 2016 (v.16) Instructions on page 1

Instructions: Use this guide to generate a regression equation using Excel. Three methods are described: creating a scatter plot, use the built-in Excel functions, and by running a regression report using statistics tools.

Data requirement: Two variables, quantitative data
Sample Data: Test sales and post-launch sales

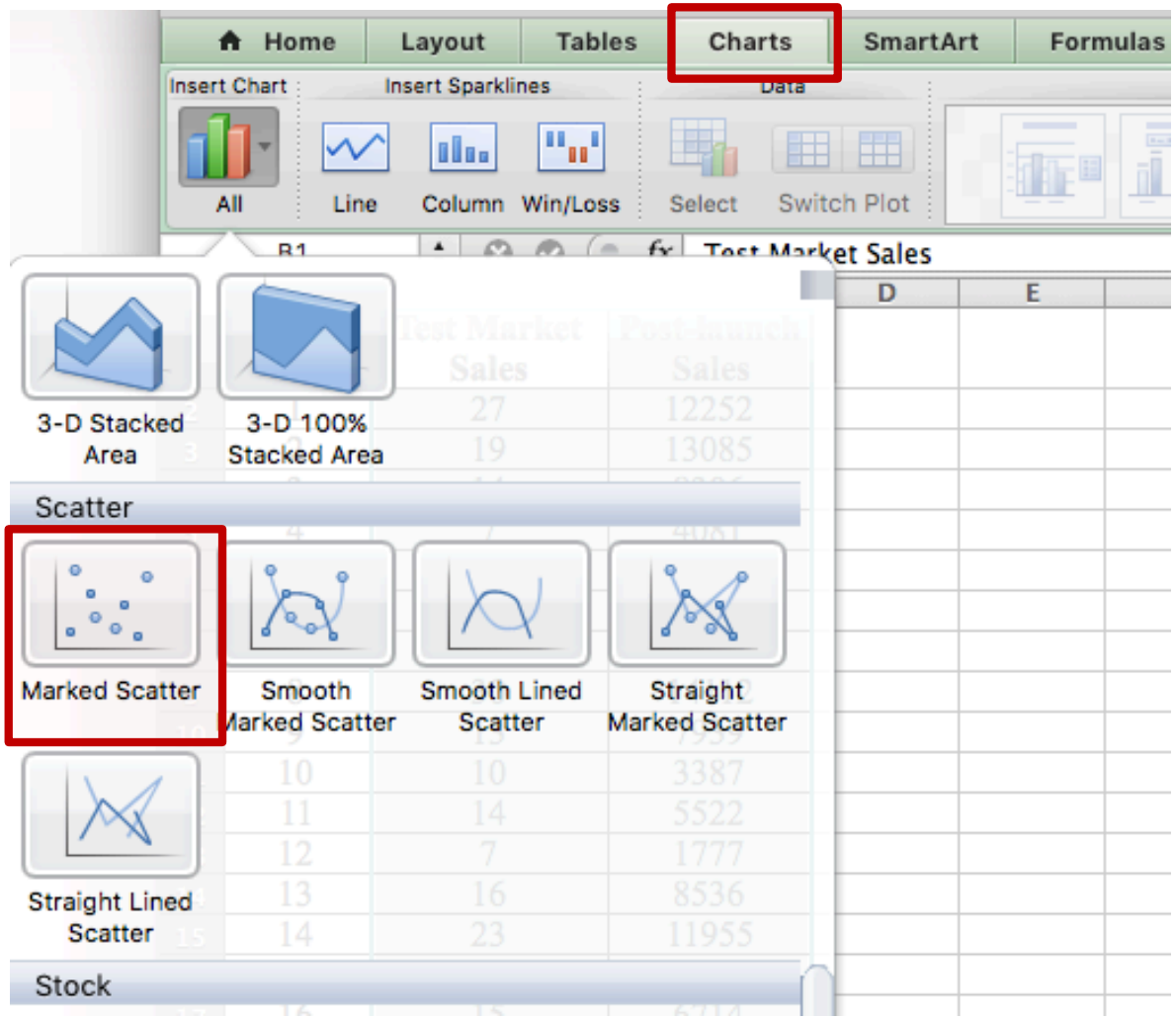
Step	Windows Instructions + Screen Shot
	Scattler Plot Method
1. Arrange the data into two adjacent columns so that the independent variable is in the left column and the dependent variable is in the right column.	

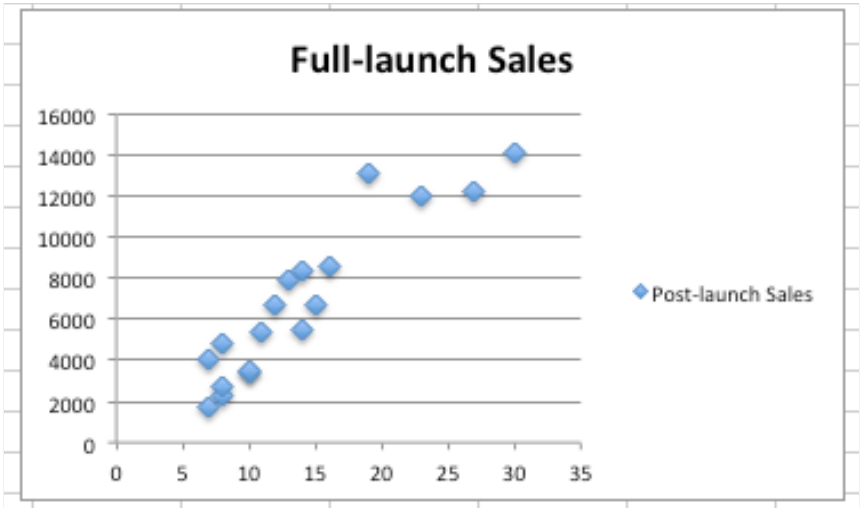
	A	B	C
1	Product	Test Market Sales	Post-launch Sales
2	1	27	12252
3	2	19	13085
4	3	14	8306
5	4	7	4081
6	5	8	4807
7	6	12	6666
8	7	11	5374
9	8	30	14112

In this example, the data are in columns B and C. The values in column B correspond to the X-axis (independent variable) and the column C data correspond to the Y-axis (dependent variable).
Note: Data for your dependent variable should always be the in right column.

2. Select your data and insert a Marked Scatter chart.

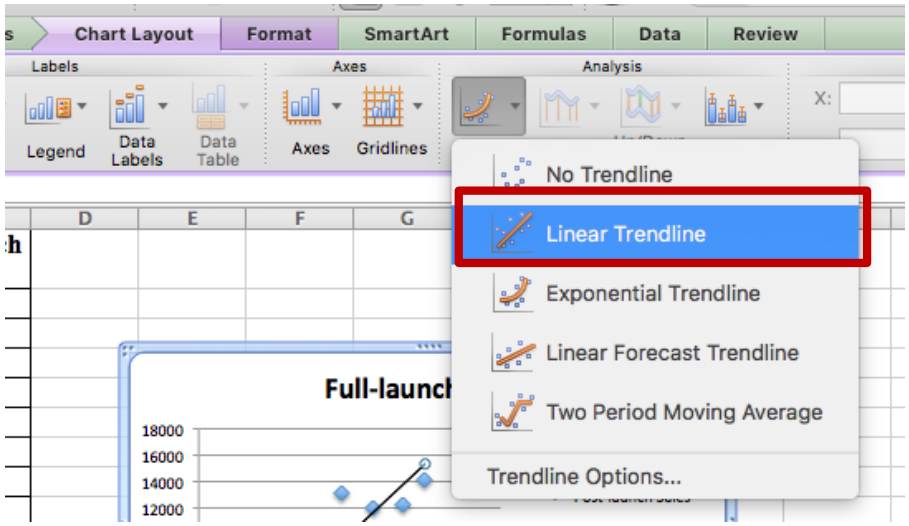
After highlighting the data in both columns, including the header rows, select the insert tab. Navigate to the Charts and select a Marked Scatter chart.





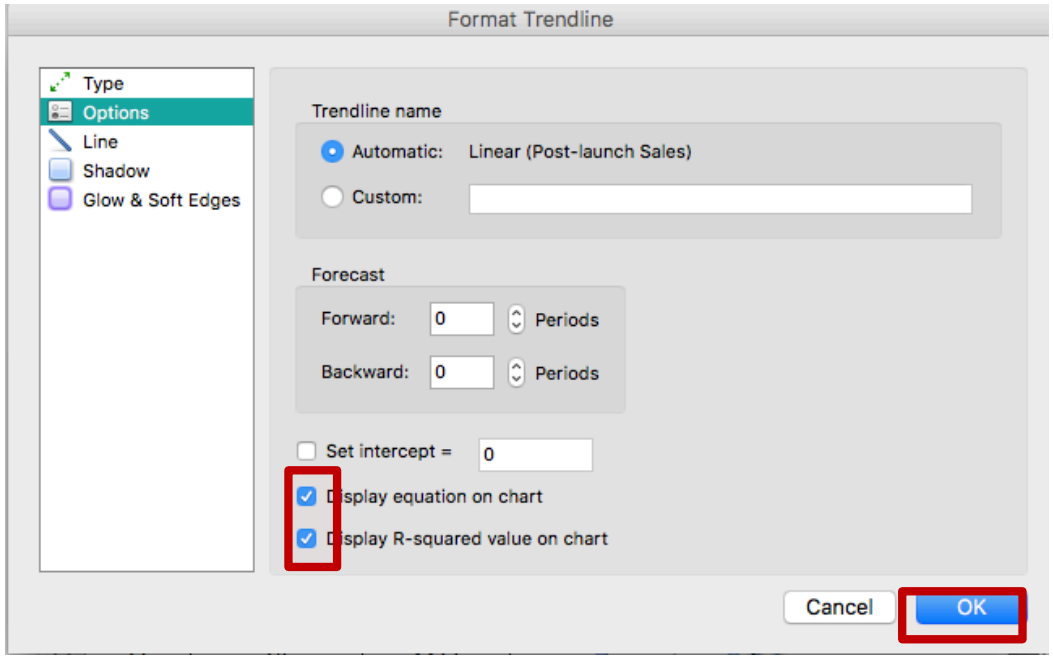
3. Display the best fit line on the scatter plot.

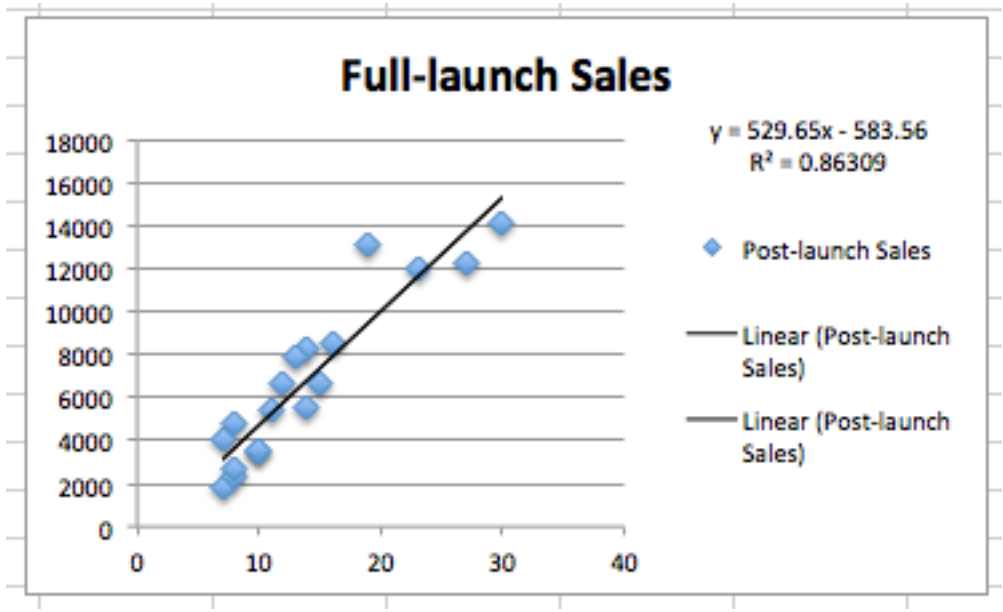
Navigate to the Chart Layout menu and add a Linear Trendline. After adding the trendline select More Options at the bottom of the menu.



4. Display the equation and R-squared value of the trendline on the chart.

Navigate to the Options tab. Select Display equation on chart and Display R-squared value on chart. Click OK.





Excel Function Method

1. Label area to calculate elements of regression equation.

Labels cells for the slope, intercept, and the R-squared value in the sheet.

	A	B	C	D	E	F
1	**Product**	**Test Market Sales**	**Post-launch Sales**			
2	1	27	12252			
3	2	19	13085		SLOPE	
4	3	14	8306		INTERCEPT	
5	4	7	4081		RSQ	
6	5	8	4807			
7	6	12	6666			
8	7	11	5374			
9	8	30	14112			
10	9	13	7939			

2. Use the SLOPE function to calculate the slope.

=SLOPE(*known y's*, *known x's*)

Type the SLOPE function into the function box. Select the column of dependent data for the first parameter of the function and select the column of independent data for the second parameter. For this example, the known y's are populated in the cell range C2:C19, so that is the range for that parameter of the formula.

F2 =SLOPE($C$2:$C$19,$B$2:$B$19)

	A	B	C	D	E	F	G
1		**Test Sales**	**Post-launch Sales**				
2		27	12252		SLOPE	529.651	
3		19	13085		INTERCEPT		
4		14	8306		RSQ		
5		7	4081				
6		8	4807				
7		12	6666				
8		11	5374				

TIP: To select a column of data

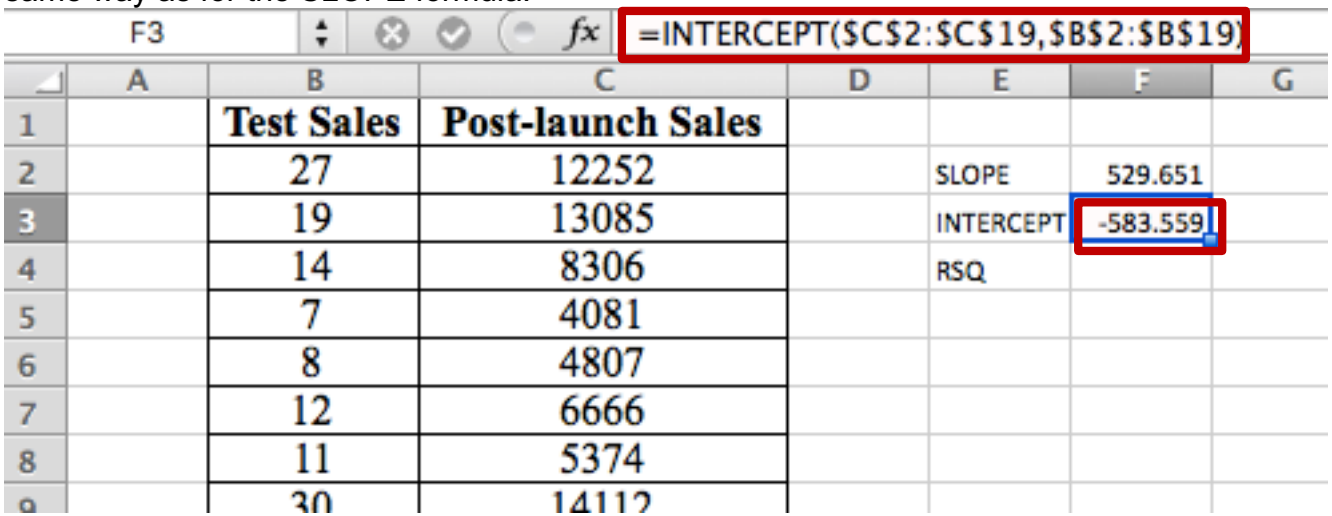
TIP: To select a block of populated cells in a column of data, select the top row in that block. The cell will be surrounded by a highlight. Now double-click in the solid square in the lower right corner of that highlight.

Another method to efficiently collect a column of data is to highlight the top cell of the column that you want and press CTRL + Shift + ↓. This will select everything between your highlighted cell and that first blank row.

3. Use the INTERCEPT formula to calculate the intercept.

=INTERCEPT(*known y's*, *known x's*)

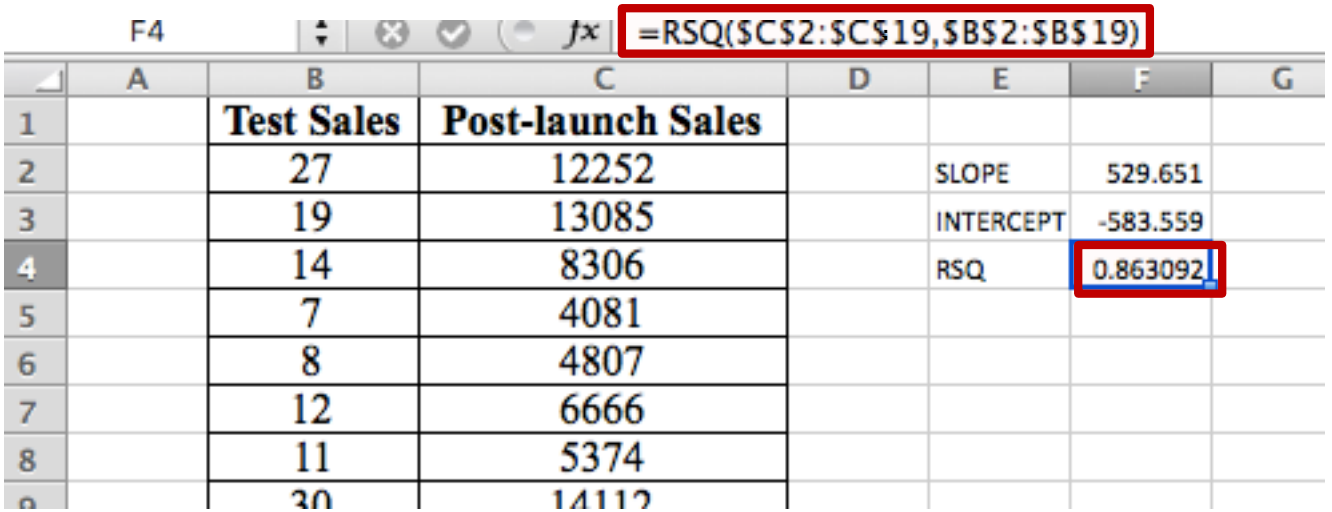
Type the INTERCEPT formula into the formula box. Select the column of dependent data for the first parameter of the formula and select the column of independent data for the second parameter the same way as for the SLOPE formula.



	A	B	C	D	E	F	G
1		Test Sales	Post-launch Sales				
2		27	12252		SLOPE	529.651	
3		19	13085		INTERCEPT	-583.559	
4		14	8306		RSQ		
5		7	4081				
6		8	4807				
7		12	6666				
8		11	5374				
9		30	14112				

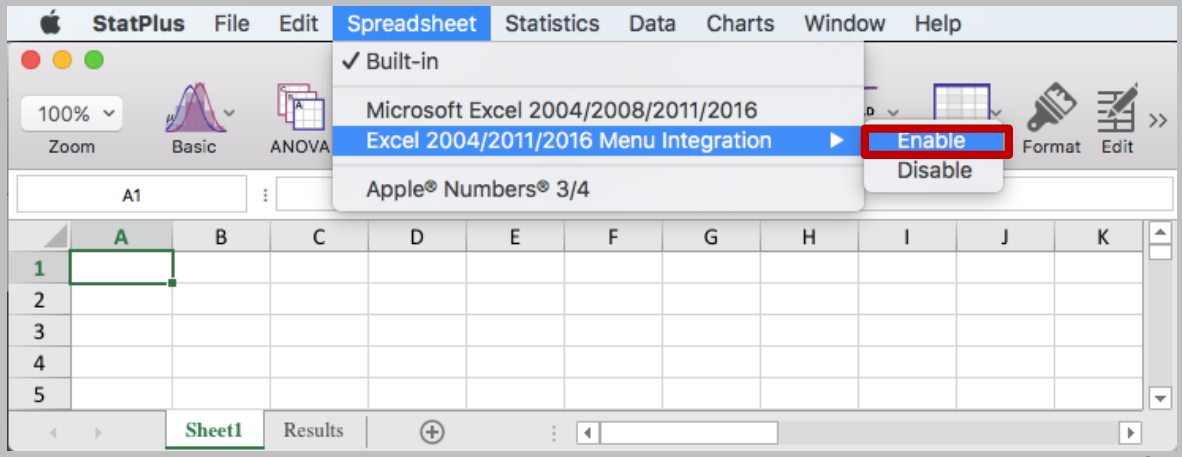
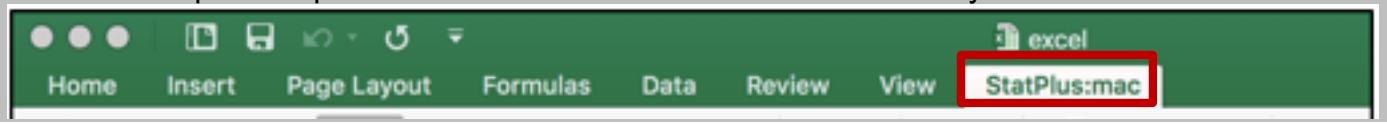
4. Use the RSQ formula to calculate the R-squared value.

=RSQ(*known y's*, *known x's*)
Type the RSQ formula into the formula box. Select the column of dependent data for the first parameter of the formula and select the column of independent data for the second parameter the same way as for the SLOPE and INTERCEPT formula.



	A	B	C	D	E	F	G
1		Test Sales	Post-launch Sales				
2		27	12252		SLOPE	529.651	
3		19	13085		INTERCEPT	-583.559	
4		14	8306		RSQ	0.863092	
5		7	4081				
6		8	4807				
7		12	6666				
8		11	5374				
9		30	14112				

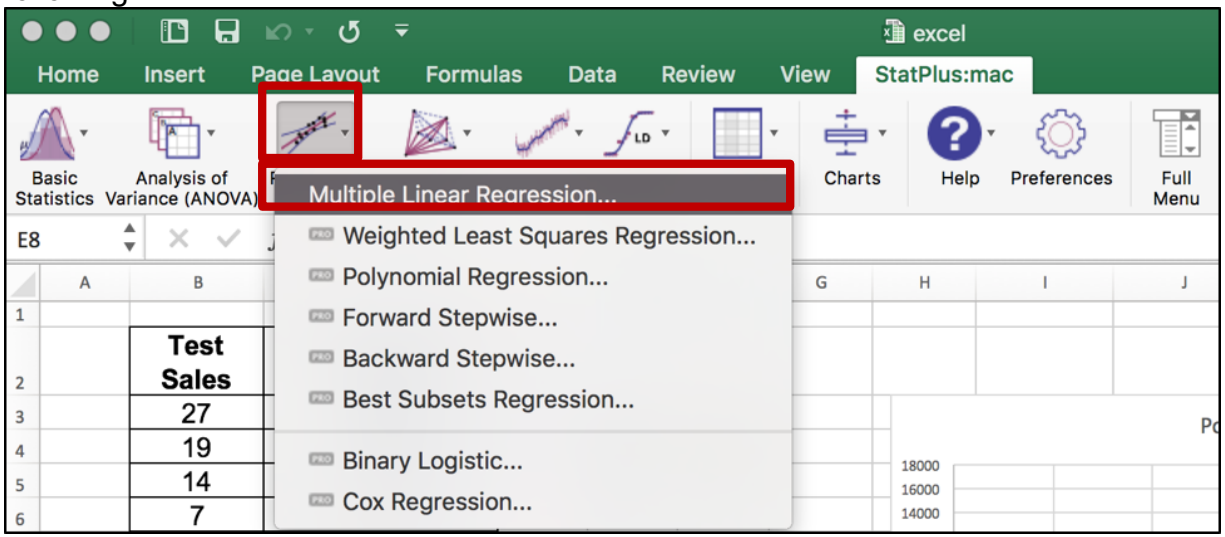
Data Analysis Method

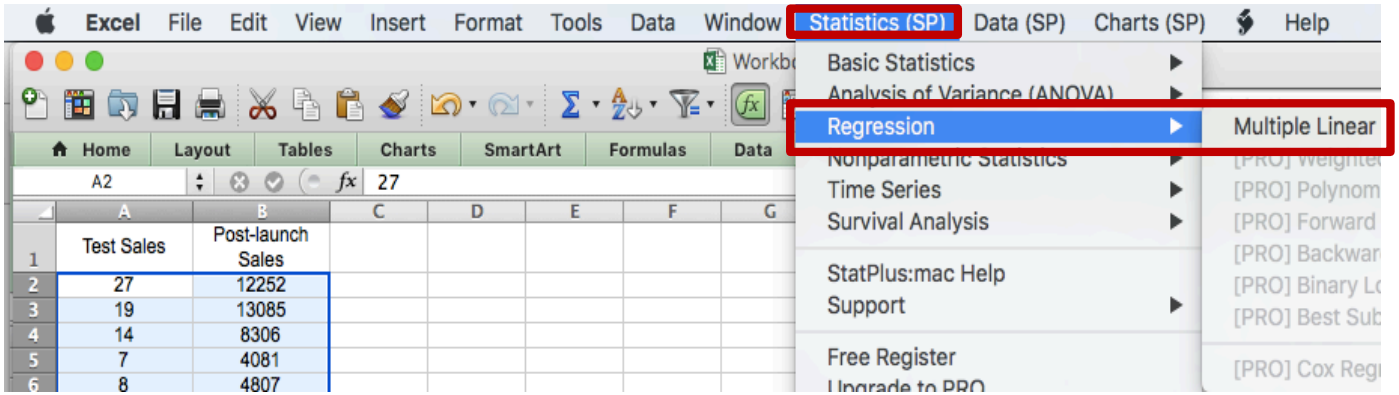
Data Analysis Tool Setup: Download StatPlus:mac to find the regression using data analysis.	Excel for Mac 2011 and earlier does not have a built-in tool pack like Windows and later Mac versions, so you'll have to download an alternative statistics package to use the data analysis method. One common free package is StatPlus:mac which can be found at https://www.analystsoft.com/en/products/statplusmacle/ Use the free download button on this webpage to begin installing the tool. Once installed, you will want to integrate the tool into you Excel menu. Open StatsPlus, and in the **Spreadsheet** menu, select Excel Menu Integration / Enable,  *Note:* This integration requires Administrator access, and you may be prompted for Administrator username and password. Once this step is complete a new StatPlus:mac tab will be added to your menu bar. 

1. Use the Data Analysis tool to insert a multiple linear regression.

Open the data analysis tool by accessing the **StatsPlus:mac** menu and selecting Regression/Multiple Linear Regression.

Note: Depending on details of your StatsPlus integration, you may see one or the other of the following:

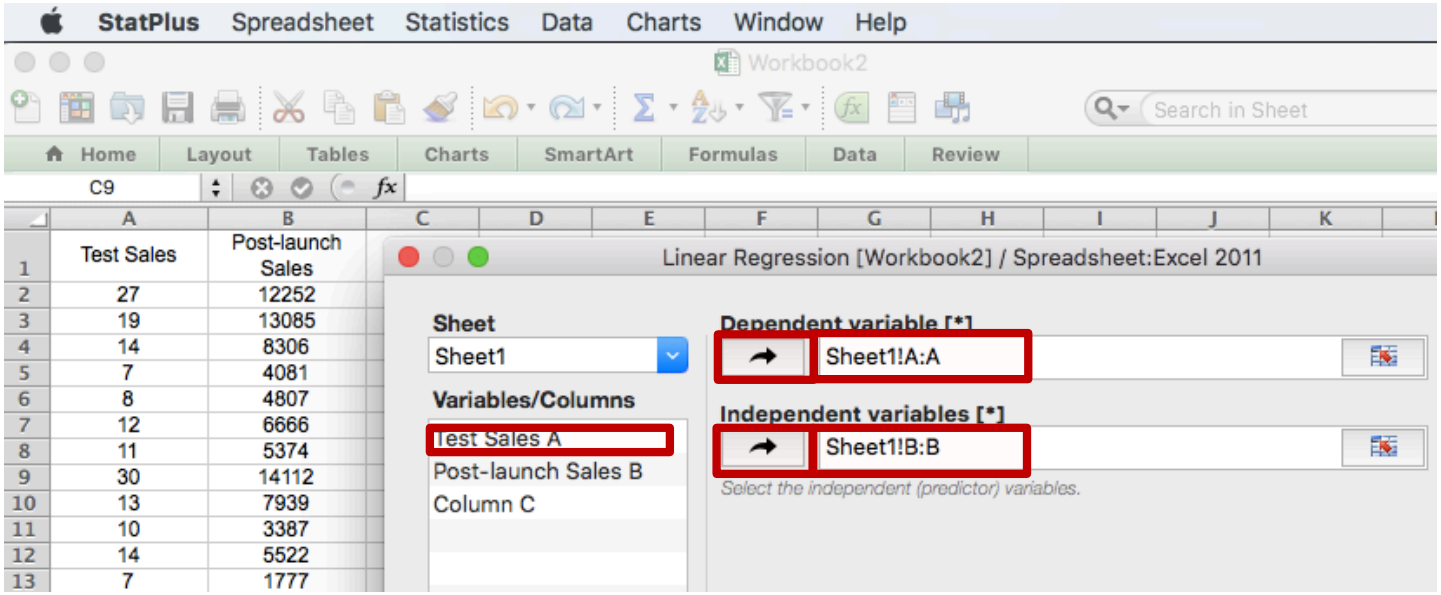




2. Enter data.

Select the dependent variable in the Variables/Columns box, then click the arrow to paste the data range in the Dependent variable field. Verify that the range selected corresponds to the dependent variable.

Repeat this process with the independent variable to paste the range in the Independent variables field.



3. View the regression table.

The R-squared value is found under Regression Statistics. The intercept and slope can be found in the bottom table.

	A	B	C	D	E	F	G	H	I	J
1	**Linear Regression**									
2										
3	**Regression Statistics**									
4										
5	*R-Squared*		0.8630							
6	*Adjusted R-Squared*		0.8545							
7	*S*		2.59667							
8	*MSE*		107.88317							
9	*RMSE*		10.38668							
10	*PRESS*		154.4555							
11	*PRESS RMSE*		2.92931							
12	*Predicted R-Squared*		0.80399							
13	*N*		18							
14	**Test Sales = 2.86764 + 0.00163 * Post-launch Sales**									
15										
16										
17	**ANOVA**									
18		*d.f.*	*SS*	*MS*	*F*	*p-value*				
19	*Regression*	1.	680.11683	680.11683	100.86716	2.58663E-8				
20	*Residual*	16.	107.88317	6.7427						
21	*Total*	17.	788.							
22										
23		*Coefficient*	*Standard Error*	*LCL*	*UCL*	*t Stat*	*p-value*	*H0 (5%)*	*VIF*	*TOL*
24	**Intercept**	2.86764	1.26619	0.18344	5.55184	2.26478	0.03777	*rejected*	**	**
25	**Post-launch Sales**	0.00163	0.00016	0.00129	0.00197	10.04326	2.58663E-8	*rejected*	**	**
26	*T (5%)*	2.11991								
27	*LCL - Lower limit of the 95% confidence interval*									
28	*UCL - Upper limit of the 95% confidence interval*									
29	*** - Requires Pro version, please upgrade.*									