Excel #12: Introduction to Regression

You have some data on purchases by some customers. You know their age and gender and you know how much money they spent at your business last year. You can estimate their personal income—either through surveys or through census bureau estimates based on their home addresses.

Age	Gender	IncomeEst	Purchases
32	M	56,000	1200
27	F	72000	2500
25	F	32000	800
31	F	96000	2800
24	M	18000	100
39	M	88000	1800
31	F	41000	1600
25	M	31578	1098
32	M	60032	1739
25	M	26350	922
23	F	23219	856
23	F	31594	1062
26	F	72848	2128
28	F	51786	1599
31	F	30479	1050
29	F	42306	1347
29	F	95138	2658
22	F	40285	1341
32	M	84062	2366
33	F	76850	2197
21	M	71009	2065
20	M	32599	1142
22	M	20888	856
35	M	46616	1457
31	F	22585	825
35	F	22908	811
21	F	75670	2169
35	M	64577	1892
25	M	83424	2394
32	F	42973	1357

You want to know how the demographic variables (age, gender, and income) affect the total purchases. Then you can target your marketing to similar groups of people. You can use statistics—particularly regression analysis—to determine the coefficients of an equation:

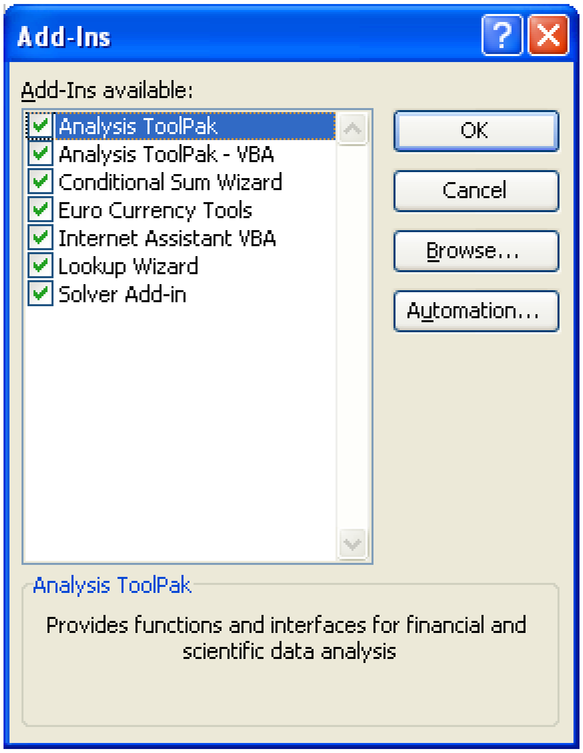
Sales = b0 + b1Age + b2Gender + b3Income

Before you can use the Excel Regression tool, you first have to convert all text data to numbers (called dummy variables). In this problem, you need to convert the gender values to numbers. It does not matter what you use, but it is easier to interpret the results if you use zero and one.

Select the data in the Gender column.
Press Ctrl+h to activate the search and replace dialog.
Enter F in the “Find what” box.
Enter 0 (zero) in the “Replace with” box.
Press Ctrl+a to replace all instances.
Enter M in the “Find what” box.
Enter 1 (one) in the “Replace with” box.
Press Ctrl+a to replace all instances.
Click the Close button.

To run the regression tool, it must first be installed. On your own computer, you might have to go back to the original installation disk and add the analytical tools if you did not select them initially. Next time, choose the option to install everything.

Even if the basic tools are installed, they might need to be loaded.
Select Tools on the main menu.
See if “Data Analysis…” appears at the bottom of the list. If it
If not, select Add-Ins from the list.
Check the Analysis ToolPak option (and the rest of them while you are at it).
Click the OK button.



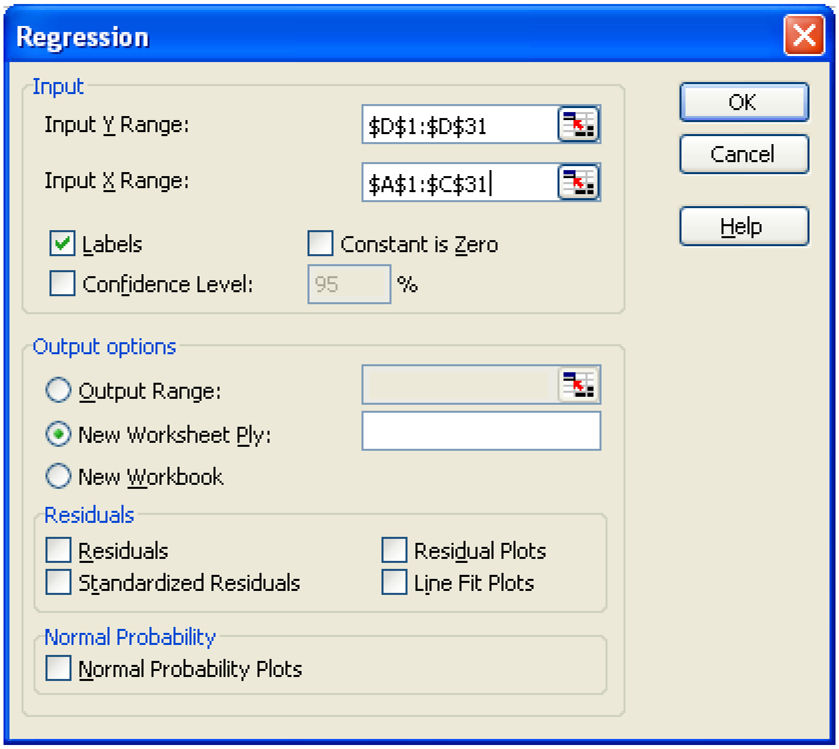
Add-Ins available:

☑ Analysis ToolPak
☑ Analysis ToolPak - VBA
☑ Conditional Sum Wizard
☑ Euro Currency Tools
☑ Internet Assistant VBA
☑ Lookup Wizard
☑ Solver Add-in

Analysis ToolPak

Provides functions and interfaces for financial and scientific data analysis

Now you can set up the regression.
Choose Tools/Data Analysis from the main menu.
Select Regression from the list and click the OK button.
Click the selection button for the “Input Y Range.”
Select the entire Purchases column, including the top label.
Press Enter.
Click the selection button for the “Input X Range.”
Select all of the rows for the other columns (Age, Gender, Income), including the top labels.
Press Enter.
Set the check mark in the Labels box.



Input

Input Y Range: $D$1:$D$31

Input X Range: $A$1:$C$31

☑ Labels □ Constant is Zero

□ Confidence Level: 95 %

Output options

○ Output Range:

● New Worksheet Ply:

○ New Workbook

Residuals

□ Residuals □ Residual Plots

□ Standardized Residuals □ Line Fit Plots

Normal Probability

□ Normal Probability Plots

Click the OK button.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.954877
R Square	0.91179
Adjusted R	0.901612
Standard E	209.1547
Observatio	30

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	3	11756696	3918899	89.58363	7.85E-14
Residual	26	1137388	43745.7		
Total	29	12894084			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	549.375	223.1861	2.461512	0.020786	90.60955	1008.141	90.60955	1008.141
Age	-10.66362	8.046898	-1.325184	0.196638	-27.20426	5.877013	-27.20426	5.877013
Gender	-164.9708	77.0954	-2.139827	0.041923	-323.4427	-6.498942	-323.4427	-6.498942
IncomeEst	0.026272	0.001646	15.9565	6.01E-15	0.022888	0.029656	0.022888	0.029656

The standardized regression results are stored in a separate worksheet.
You can also put them in the main worksheet, but be careful that you do not overwrite the data.

The coefficient values are displayed in the bottom table, so your equation becomes

Purchases = 549 – 10.7*Age – 165*Male + 0.026*Income

The P-Value tells you whether the coefficient is significantly different from zero. All except the Age coefficient have values less than 0.05, which makes them significant.

The R Square value indicates that 91 percent of the observed variation is explained by these variables; which indicates that the equation is good. (It would likely be even better if you had more rows of data.)

You can use the estimated equation to forecast new values. Simply plug in values for the variables and perform the calculation.

You can also use it to analyze your sales. The (possibly) negative sign on Age indicates that your business sells more to younger customers. The negative sign on Male (Gender) indicates that women are spending more money than men. The positive sign on Income indicates that people with more income (regardless of age and gender) will spend more at your business.