Simple Regression in MS Excel
Instructions

(1) If you do not see *Data Analysis* in the menu, this means you need to use *Add-ins*, and make sure that the box in front of Analysis ToolPak is checked. Instructions for using *Add-ins* are in *Assignments*.

(2) Go to *Data Analysis—Regression* in menu.

(3) You will have to indicate where the Y-variable (dependent variable) is located, e.g., d4:d13; you will also have to indicate where the X-variable (independent variable) is located, e.g., b4:b13.

Input Y Range: d4:d13
Input X Range: b4:b13

Optional:
(a) Check the box in front of *Confidence Level* 95% (95% is the default so you do not have to change this). This will give you a 95% confidence interval for the b_o and b_1 coefficients. If you do not check this, the program will run but you will not get the confidence interval.
(b) If you would like to see the residuals (also known as deviations), check the box in front of *Residuals*. [You can also do *Residual Plots*.]

You have to indicate where you want the regression output to appear. You will probably want the output to appear either on the same page or on another worksheet. If you want the output to appear on the same page, then check the circle in front of *Output Range* and indicate where the output should go. If the data appear in rows, say, b4: b13 and d4:d13, then your output should not appear in the first 13 rows. I would indicate a14 next to *Output Range*. Of course, you can check the circle in front of *New Worksheet Ply* and your output will appear on another worksheet. This may be a good idea if you are afraid that the output is too large to appear on the same page as the input.

Example: A bookstore wants to determine whether there is a relationship between shelf space and number of books sold.

Shelf Space (in feet) (X)	Number of Books Sold
1	14
2	18
3	22
4	25
5	30
6	34
7	40
8	43
9	50
10	54

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.997115226
R Square	0.994238773
Adjusted R Square	0.99351862
Standard Error	1.09336844
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1650.436364	1650.436364	1380.593156	3.0194E-10
Residual	8	9.563636364	1.195454545		
Total	9	1660			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.4	0.746912838	11.24629216	3.50917E-06	6.677614793	10.12238521
X Variable 1	4.472727273	0.120375903	37.15633399	3.0194E-10	4.195139762	4.750314783

This regression is very significant; the F-value is 1380.59. If the X-variable explains very little of the Y-variable, you should get an F-value that is 1 or less. In this case, the explained variation (due to regression = explained by the X-variable) is 1,380.59 times greater than the unexplained (residual) variation. The probability of getting the sample evidence (the X and Y input data) if the X and Y are unrelated (that is the Ho) is .00000000030194. In other words, it is almost impossible to get this kind of data as a result of chance.

The intercept coefficient (Y-intercept) is the bo; in the above problem it is 8.4.
The X Variable 1 coefficient (slope term) is the b1; in the above problem it is 4.47 (rounded).

The regression equation is:
Sales = 8.4 +4.47 (shelf space).
In theory, if no shelf space is assigned to the book (book must be ordered from catalog), you will sell 8.4 copies. Every foot of shelf space will increase sales by 4.47 books.

The correlation coefficient is .997. It is almost a perfect 1. To get 1, all the points would have to be on a line and all the residuals (deviations) would be 0.
The coefficient of determination, r^2, is 99.4%. There is very little unexplained variation (in fact, the MSE, mean square error is only 1.195).

Another way to test the regression for significance is to test the b_1 term (slope term which shows the effect of X on Y). This is done via a t-test. The t-value is 37.156 and this is very, very significant. The probability of getting a b_1 of this magnitude if Ho is true (the null hypothesis for this test is that $B_1 = 0$, i.e., the X variable has no effect on Y) is 3.0194E-10 or .00000000030194.

Note that this is the same sig. level we got before for the F-test. Indeed, the two tests give exactly the same results.