**Linear Regression by Hand and in Excel**

Get the data – 12 Month Marketing Budget and Sales



X - Spend, Y - Sales

**Simple Linear Regression Math by Hand**

1) Calculate average of X variable Avg(X).

2) Calculate difference between Avg(X) and X i.e., Avg(X) - X.

3) Square the differences and add all the squares. ∑(Avg(X) - X)^2. Label it as SSxx.

4) Calculate average of Y variable Avg(Y).

5) Calculate difference between Avg(Y) and Y i.e., Avg(Y) - Y.

6) Multiply Avg(X)-X and Avg(Y)-Y and add all the products. ∑(Avg(X) - X) \* (Avg(Y) - Y). Label it as SSxy.

**SSxy - Covariance of X and Y**

**SSxx - Variance in X**

Slope = SSxy / SSxx = 10.6221

Y Intercept = Avg(Y) - Slope \* Avg(X) = 1383.4713

Now Linear Equation is Y = 10.6221 \* X + 1383.4713

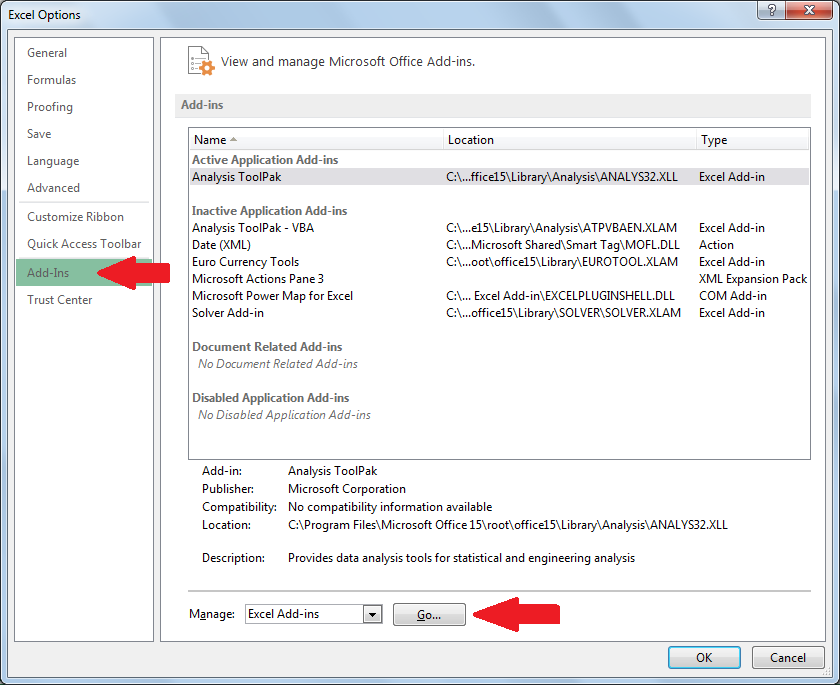


**Simple and Multiple Regression with Excel’s Data Analysis Tools**

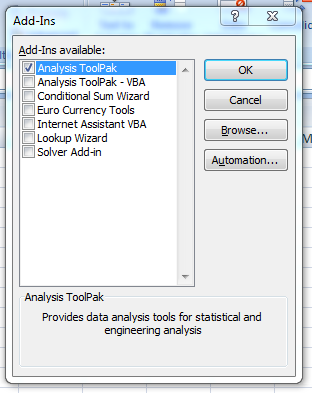
1) Go to Excel File Tab or Excel office button.

2) Click on Excel Options and go to Add-Ins.

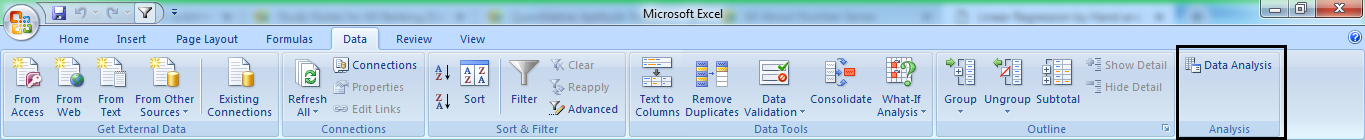
3) Go to Manage, select Excel Add-Ins and click on Go.



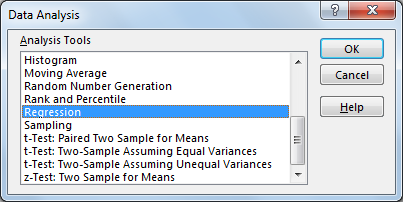
4) Check the Analysis ToolPak and click on OK. It installs Excel Data Analysis Tool.



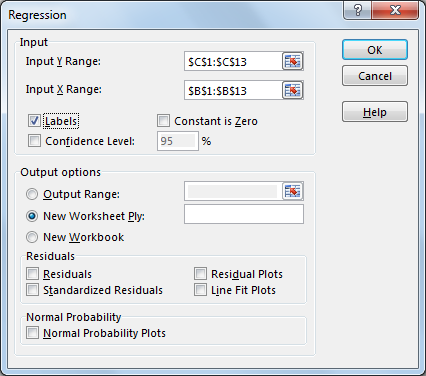
5) Go to Data Tab and on right corner we can see Analysis Tool as "Data Analysis".



6) To use click on Data Analysis button, and we can see different built-in analyses.



7) Launch Regression tool



8) Input Y Range is where the response variable (Sales in our case) is located.

9) Input X Range is the range of predictor variables (Spend).

* If there were additional X variables, they would all have to be next to each other.  No gaps between X variables allowed.

10) Labels being checked means you have a header at the top of your X and Y range.

* Confidence Level – Adds another confidence interval at selected confidence level.
* Constant is Zero – Forces the X coefficient to capture more of the error.
  + Almost no reason to ever use this option unless your data has a theoretical reason to pass through the origin.
  + The regression equation is fundamentally changed as well ([PDF Notes](http://courses.washington.edu/qsci483/Lectures/20.pdf))
* Residuals – For every row, it provides the error / difference between predicted and actual values.
  + Standardized Residuals is normalized with mean zero and standard deviation of one.
  + Residual Plots charts the residuals by each variable.
  + Line Fit Plot charts the predicted results and the actual results by each variable
* Normal Probability Plots – Checks normality of your data.  Should see something close to a straight line.