How to Perform Regression Analysis

Please see the following video demonstration: Regression Video Guide

Adding the Analysis Toolpak to Excel

The Analysis Toolpak is a Microsoft Excel add-in program that needs to be loaded into **Excel 2013 or Excel 2016.**

1. From **Excel 2013 or Excel 2016**, click the **File** tab, and then click **Options**.
2. Click **Add-Ins** and in the **Manage** box, select **Excel Add-ins**.
3. Click **Go… button**.
4. In the **Add-Ins available**: box, select the **Analysis ToolPak** check box, and then click **OK**.
5. After you load the Analysis ToolPak, the **Data Analysis** command is available in the **Analysis group** on the **Data tab**.

Creating a Graph and Best Fit Formula

1. Have your x and y data side-by-side
2. Select all, including the labels
3. At the top of Excel, click insert and then choose 'scatterplot'
4. After the scatterplot appears, right click on a plot in your graph and select 'add trendline'
5. Then in the box to the right, select 'display your equation on chart'
   Investigate and choose one of the following topics to be the focus of your research and analysis

Performing Regression Analysis

1. Highlight all the data
2. Click **Data** at the top, then **Data Analysis** on the right side to find **Regression Analysis** on the list
3. Click **Regression Analysis**
4. Click the box to allow labels
5. Ensure your X and Y data are correctly chosen
6. Your regression analysis should then appear!

You may now try and practice this using the **Simple Data.csv** dataset.

Interpreting Regression Analysis

Multiple R: It measures the strength of association between the independent (explanatory) variables and the dependent variable (the variable we wish to forecast). Its value varies between 0 and 1; the higher value, the stronger the association.

R^2: R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model. If the R^2 is 0.75, then approximately 75% of the observed variation can be explained by the regression model's inputs.

Adjusted R^2: Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases when the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected. Typically, the adjusted R-squared is positive, not negative. It is always lower than the R-squared.

Observations: simply the # of data entries

ANOVA section: these numbers are used to get the above results

Intercept: first data nearest y axis tells what is the value they *can* get with that x value; Using $y= B_0 + B_1x$, you can put in values given to determine the formula

Standard Error: average distance observed values fall from the regression line; tells how "wrong" the regression model is on average using the units of the response variable

t-stat: ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error. >2 or <2 is acceptable.

p-value: Is the relationship significant between 2 variables? If this is large, the significance cannot be trusted, and a variable may have to be eliminated. <0.05 is good (95% confidence interval)

Lower and Upper 95% Confidence Levels: 95% of the data falls in this range

Use of y=mx+b: Coefficients in the graph can be used to predict any new variable; TRY IT!

Interpretation example with a low R^2 and low p-values: Link