Linear Regression Analysis

You are required to follow these steps for doing and presenting Linear Regression Analysis

Step # 01: Selection of Variables

In this step you are required to identify the specific factors[1] (independent variables) which affect the profitability[2] (dependent variable) of the firm in which you have done your Internship. For the purpose of identification you must study research articles relevant to **Profitability Determinants**. You should look at financial statements of the firm and identify the variables that are likely to have an impact on firm's profitability (*You are also advised to seek guidance from the instructor before finalizing the variables*).

After the selection, you will provide brief introduction of the selected variables (dependent as well as independent variables) along with the sources from where the data has been collected. You are required to collect financial data of the firm for recent FIVE years or more and test for factors affecting a firms' profitability. Also find evidence that out of the selected Independent variables which variable is/are significant in determining the profitability by using **Multiple Linear Regression Analysis.**

This step should be presented in the following way

Dependent Variable	Independent Variables	Year Under Consideration	Analysis Technique	Data Analysis Software	Data Collection Source
Profitability measure e.g. ROA, ROE	Total Assets, Sales, Equity etc.*	20XX – 20YY			
*Make sure that you have to take more than one independent variable for analysis.					

Step # 02: Model Specification

In this step you will be required to present your model in mathematical equation form.

$$ROA = \alpha + \beta_1 TA + \beta_2 S + \beta_3 EQ + \varepsilon$$

Where:

ROA = Return on Assets

TA= Total Assets

S = Total sale of the firm

EQ= Total equity of the firm

ε = Error/disturbance/residual

---

[1] Total Asset, Total Equity, Total Liability, Firm size, Sales etc.

[2] Profitability can be measured by Return on Assets (ROA) or Return on Equity (ROE) or any other measure of your choice.

Step # 03: Data Analysis

In this step are required to show the output results of the linear regression. Procedure of how to run a Linear Regression in Excel is given in later part of the document.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.5173425
R Square	0.2676432
Adjusted R Square	0.0235243
Standard Error	2.217208
Observations	9

ANOVA

	df	SS	MS	F	Significance F
Regression	2	10.779481	5.3897409	1.0963644	0.3927968
Residual	6	29.496073	4.916012		
Total	8	40.275555			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.6181742	1.9731876	0.8200812	0.4435207	-3.2100412	6.4463904	-3.210041	6.4460468
Total Assets	0.2159459	0.3262453	0.6619129	0.5326100	-0.5823475	1.0142395	-0.582347	1.0142335
Sales	0.3276435	0.3317540	0.9876097	0.361481	-0.4841294	1.1394165	-0.484129	1.1394165

Step # 04: Interpretation of the Results

In this step you are required to determine the significant factors that have an impact on firm's profitability and also identify their impact level. So, these findings should be based on the interpretation of:

1. Intercept Coefficient
2. Slope Coefficients along with direction of relationship (being positive/negative)
3. R square
4. Model Significance (as per **Prob (F-statistic)**)

1. **a** = 1.618 (Intercept Coefficient)

2. b_1 = 0.216 (Coefficient of Total Assets i.e. slope)
   b_2 = 0.327 (Coefficient of Sales i.e. slope)

   So our regression equation is ROA = 1.618 + 0.216(TA) + 0.327(S)

   b_1 = 0.216 means that one unit increase in TA brings 0.216 unit increase in ROA, on average.

   Note: In the above mentioned way, you are required to interpret all the slope coefficients.

   Also in the regression statistics output gives the goodness of fit measure i.e. R square.

3. **R square** = 0.267 (This means 26.7% of ROA is determined by Total Assets and Sales and the rest is unexplained due to the factors/variables not considered in the study).

4. **Model Significance (as per probability of F-statistics)** = 1.096 (0.392) means that model/results are not significant as p-value of F '0.392' is greater than 0.05(significance level). So we can infer that there is no signification relationship among the dependent and independent variables.

Important Notes

- Regression analysis based on the financial data **older or less** than recent five years will **NOT** be accepted.
- You are required to submit you the **Raw Data (excel file)** which contains the data of variables along with your internship report.


Non Availability of Financial Statements for Linear Regression Analysis

All those students who will be facing difficulty in conducting linear regression analysis due to non- availability financial statements of the organization's in which they have done their internship are required to discuss this issue with their Course Instructor via email at: Ecoi619@vu.edu.pk. They will be given guidance regarding this issue; as regression analysis is a very important part of internship report and skipping this part will **NOT** be accepted at all.
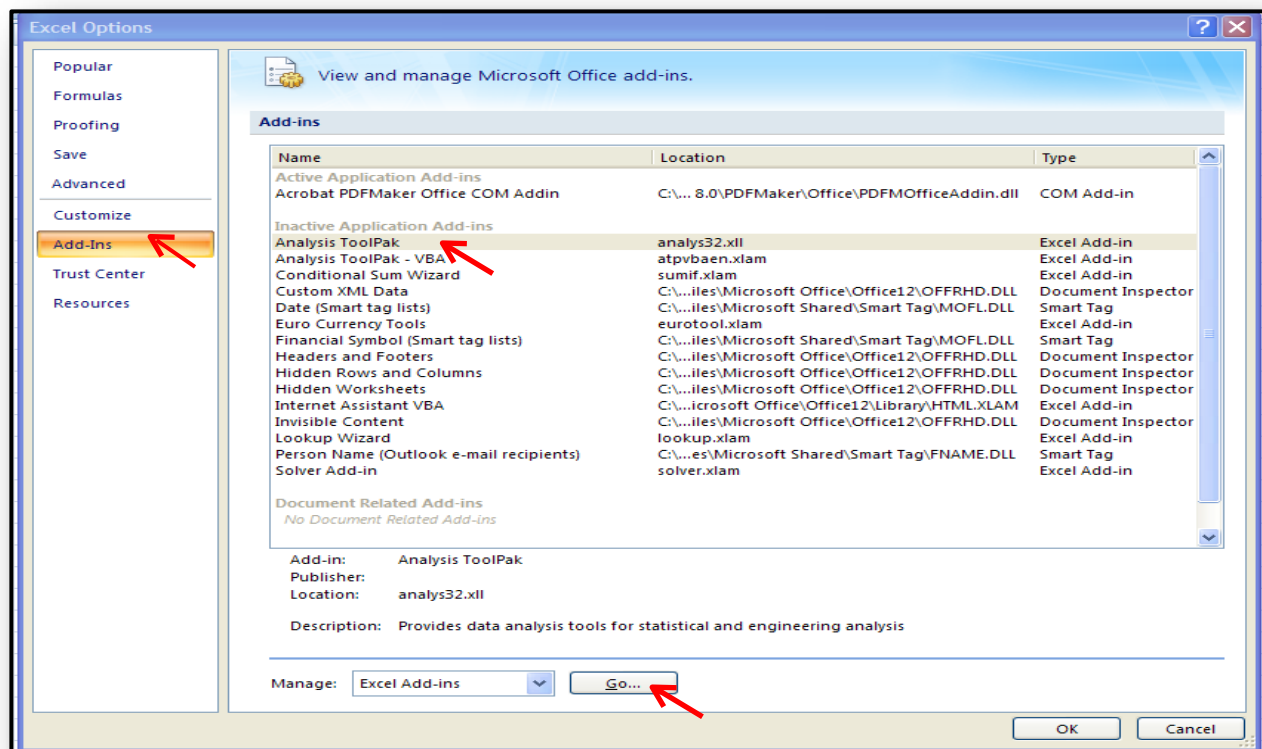
Linear Regression using Excel

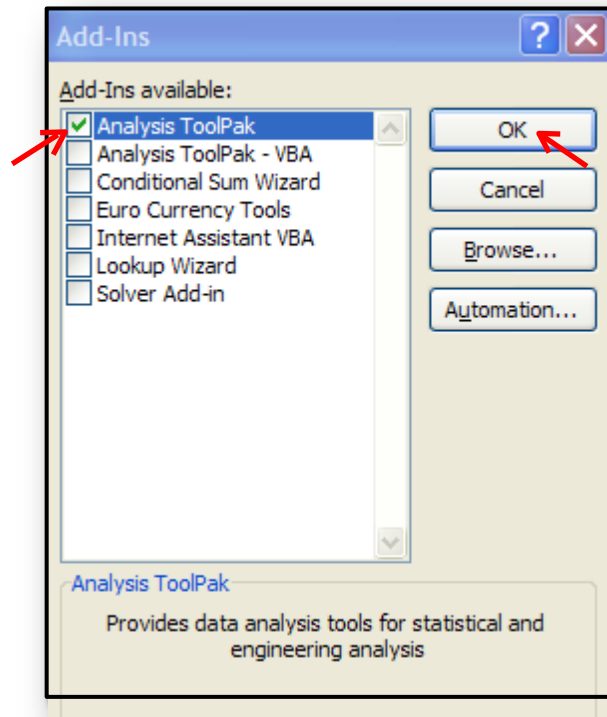## Installing the Analysis ToolPak

1. Open Microsoft Excel, click the File menu and select **Options**
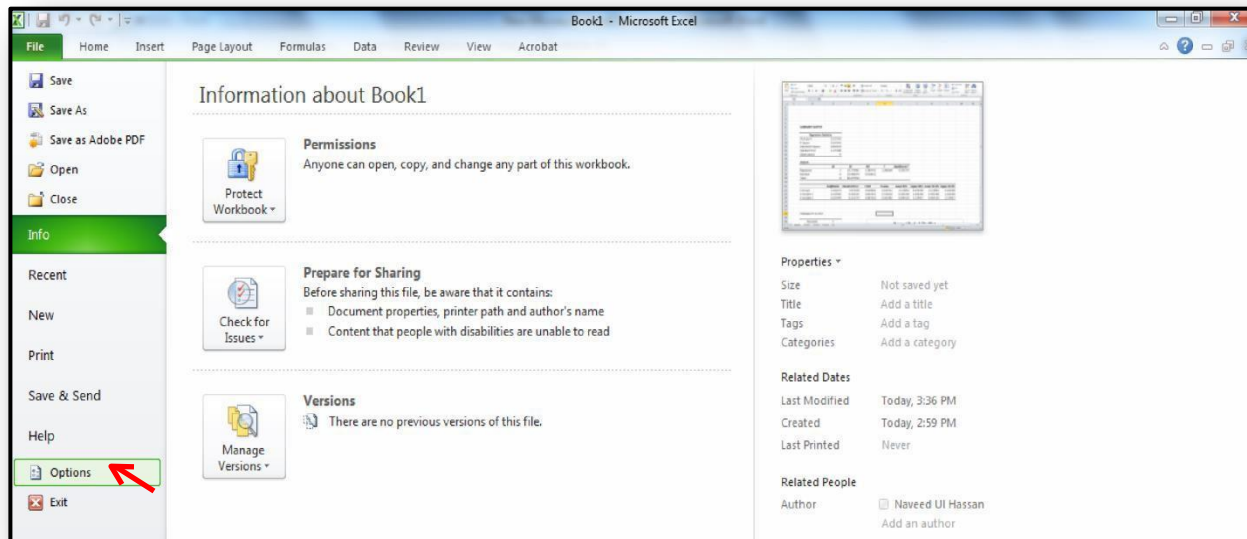
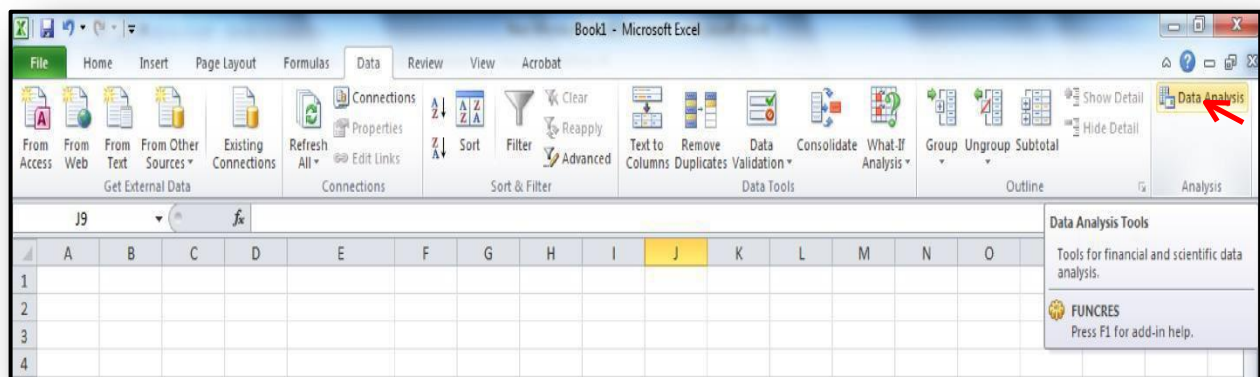The following window will open.

2. In the Microsoft Office button, go to excel options to click Add-ins
3. In the Add-Ins box, select Analysis ToolPak and click Go...

4. In the Add-Ins available box, check the Analysis ToolPak and then click OK.
5. If Analysis ToolPak is not listed in the Add-Ins available box, click Browse to locate it.

Linear Regression using the Data Analysis Add-In

After installation of Analysis ToolPak you will find Data analysis button on the right corner of data menu. Click this button

Following window will open.

1. In the data analysis tool select the regression and then click Ok.

Suppose we want to determine whether X is a function of Y and Z.

$$X_i = a + b_1Y_i + b_2Z_i + error_i$$
Where;

X_i = value of X for observation i
a = mean value of X (dependent variable) when Y and Z (independent Variable) is zero (intercept coefficient)
b = average change in X given a one unit change in Y and Z, i.e. (slope of Y and Z)
$\quad$ b_1 = value of Y for observation i
$\quad$ b_2 = value of Z for observation i

2. After clicking the regression, a regression box appears.
3. Select the Inputs for Y range (Dependent Variable) in first box.
4. Select the Inputs for X range (Independent Variable) in 2nd box.
4. Select the place where you want your output
5. Check the Labels and then click Ok

SUMMARY OUTPUT

Regression Statistics

Regression Statistics	
Multiple R	0.517342518
R Square	0.267643281
Adjusted R Square	0.023524374
Standard Error	2.21720822
Observations	9

ANOVA

	df	SS	MS	F	Significance F
Regression	2	10.77948181	5.389741	1.096364	0.392796864
Residual	6	29.49607374	4.916012		
Total	8	40.27555556			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.618174259	1.973187635	0.820081	0.443521	-3.21004195	6.446390468	-3.21004195	6.44639047
Y	0.215945998	0.326245319	0.661913	0.53261	-0.582347538	1.014239535	-0.582347538	1.01423953
Z	0.327643553	0.331754071	0.98761	0.361482	-0.484129415	1.139416521	-0.484129415	1.13941652

The output is given in the coefficients column in the last set of output

1. Intercept Coefficient
2. Slope Coefficients
3. R square
4. Model Significance (as per **Prob (F-statistic)**)