EXPERIMENT No.: 5

NAME OF EXPERIMENT: Perform Multiple Linear Regression using Data Analysis Toolbox of Excel or with Python and Interpret the regression table.

Name- Pratik Rajesh Jade

Roll no- A72

Date of Performance: _____ Sign of Teacher:

Aim: Perform Multiple Linear Regression using Data Analysis Toolbox of Excel and Interpret the regression table

Multiple linear regression (MLR), also known simply as **multiple regression**, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. **Multiple regression** is an extension of **linear** (OLS) **regression** that uses just one explanatory variable

Regression models are used to describe relationships between variables by fitting a line to the observed data. Regression allows you to estimate how a dependent variable changes as the independent variable(s) change.

Multiple linear regression is used to estimate the relationship between **two or more independent variables** and **one dependent variable**. You can use multiple linear regression when you want to know:

Formula and Calculation of Multiple Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals)

1. How strong the relationship is between two or more independent variables and one dependent variable (e.g. how rainfall, temperature, and amount of fertilizer added affect crop growth).

2. The value of the dependent variable at a certain value of the independent variables (e.g. the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition)

Steps:

1. **Launch Excel**. To begin your **multivariate** analysis in **Excel**, **launch** the Microsoft **Excel**. ...
2. Click on options. On the left side of the dialog box is a list with options. ...
3. Check the box. ...
4. Performing the **Regression**. ...
5. Data tab. ...
6. **Regression**. ...
7. Dependent Variable. ...
8. Independent Variable.

When should we use multiple linear regressions?

Multiple linear regressions is **used** to estimate the relationship between **two** or more independent variables and **one** dependent variable.

INPUT:

Roll No.	Name of Student	Department	Physics	Chemistry	BIO	CS	MATH	Total	Result
1	ABHISHEK DESHMUKH	AI	60	70	80	85	65	360	72
2	ADITYA MANOJ PATHARKAR	AI	65	59	76	86	66	352	70.4
3	ANUJ SHARMA	AI	63	62	73	75	69	342	68.4
4	ASHISH GUPTA	AI	69	65	68	69	68	339	67.8
5	KALYANE VIDWAN BALIRAJE	AI	75	69	70	65	76	355	71
6	MAHESH RAMESH BHALERAO	DS	65	68	59	63	78	333	66.6
7	DEVASHRI NANDKUMAR	DS	66	76	62	69	65	338	67.6
8	SANKET VINOD	DS	69	78	65	92	68	372	74.4
9	CHAUDHARY SHUBHAM	DS	68	75	68	90	76	377	75.4
10	VISHWANATH CHOLE	DS	76	79	76	79	78	388	77.6
11	HARSHAWARDHAN	DS	78	81	78	81	75	393	78.6
12	EKLAVYA VILAS KIROTE	DS	75	65	75	82	69	366	73.2
13	RUTIK RAMDAS GAWALI	AI	79	63	69	86	86	383	76.6
14	SNEHA ASHOK GAWARE	AI	81	69	75	80	80	385	77
15	TANMAY ANIL GORKAR	AI	82	75	65	86	86	394	78.8
16	ONKAR HANMANTRAO HAVARGE	AI	86	65	66	87	87	391	78.2
17	SUYOG SANJAY IGAVE	AI	80	66	80	85	60	371	74.2

OUTPUT:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	4.9E-15
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	276.6375	55.3275	2.3E+30	5.8178E-150
Residual	10	2.4E-28	2.4E-29		
Total	15	276.6375			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.95E-14	2.7E-14	0.724572	0.485322	-4.05475E-14	7.96E-14	-4.1E-14	7.96E-14
60	0.2	2.65E-16	7.55E+14	4.1E-145	0.2	0.2	0.2	0.2
70	0.2	1.93E-16	1.04E+15	1.7E-146	0.2	0.2	0.2	0.2
80	0.2	2.73E-16	7.33E+14	5.5E-145	0.2	0.2	0.2	0.2
85	0.2	1.56E-16	1.28E+15	2E-147	0.2	0.2	0.2	0.2
65	0.2	2.29E-16	8.74E+14	9.4E-146	0.2	0.2	0.2	0.2

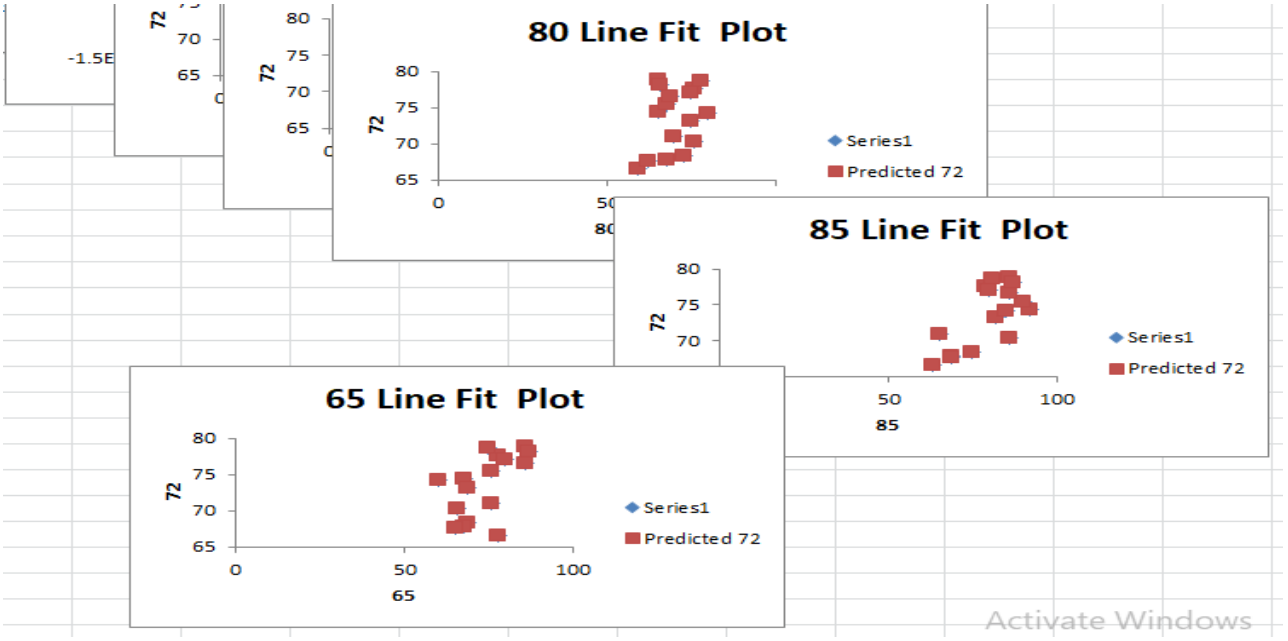
SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	4.9E-15
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	276.6375	55.3275	2.3E+30	5.8178E-150
Residual	10	2.4E-28	2.4E-29		
Total	15	276.6375			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	1.95E-14	2.7E-14	0.724572	0.485322	-4.05475E-14	7.96E-14	-4.1E-14	7.96E-14
60	0.2	2.65E-16	7.55E+14	4.1E-145	0.2	0.2	0.2	0.2
70	0.2	1.93E-16	1.04E+15	1.7E-146	0.2	0.2	0.2	0.2
80	0.2	2.73E-16	7.33E+14	5.5E-145	0.2	0.2	0.2	0.2
85	0.2	1.56E-16	1.28E+15	2E-147	0.2	0.2	0.2	0.2
65	0.2	2.29E-16	8.74E+14	9.4E-146	0.2	0.2	0.2	0.2


80 Line Fit Plot
72
80
75
70
65
0
50
80
Series1
Predicted 72
85 Line Fit Plot
72
80
75
70
50
100
85
Series1
Predicted 72
65 Line Fit Plot
72
80
75
70
65
0
50
100
65
Series1
Predicted 72
Activate Windows