Name- Pratik Jade Roll no:- A72

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?

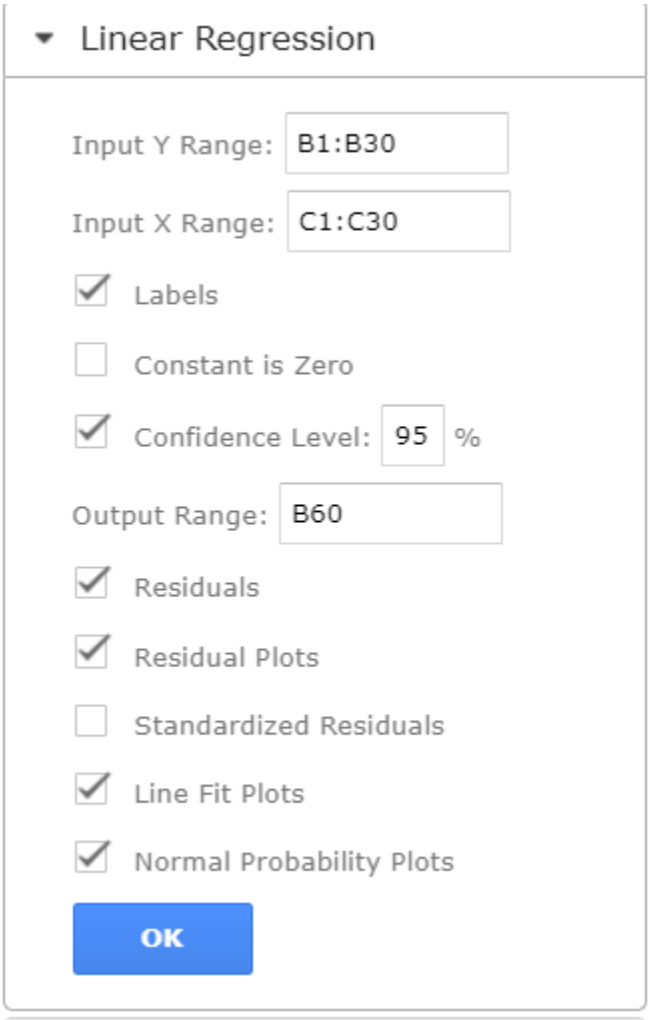
Multiple linear regressions is used to estimate the relationship between two or more independent variables and one dependent variable.

Input:-

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1

3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1
10	125	70	26	115	31.1	0.205	41	1
7	147	76	0	0	39.4	0.257	43	1
1	97	66	15	140	23.2	0.487	22	0
13	145	82	19	110	22.2	0.245	57	0

Tool:-



Result(Output) :-

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.2433882635
R Square	0.05923784683
Adjusted R Squa	0.02439480412
Standard Error	32.86499492
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1836.328313	1836.328313	1.700134151	0.2032782116
Residual	27	29162.91307	1080.107891		
Total	28	30999.24138			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Upper 95%*
Intercept	107.5894041	18.61310279	5.78030462	0.0000037818086	69.3984722	145.780336	69.3984722	145.780336
BloodPressure	0.3385474957	0.2596438319	1.303891925	0.203278209	-0.1941976359	0.8712926273	-0.1941976359	0.8712926273

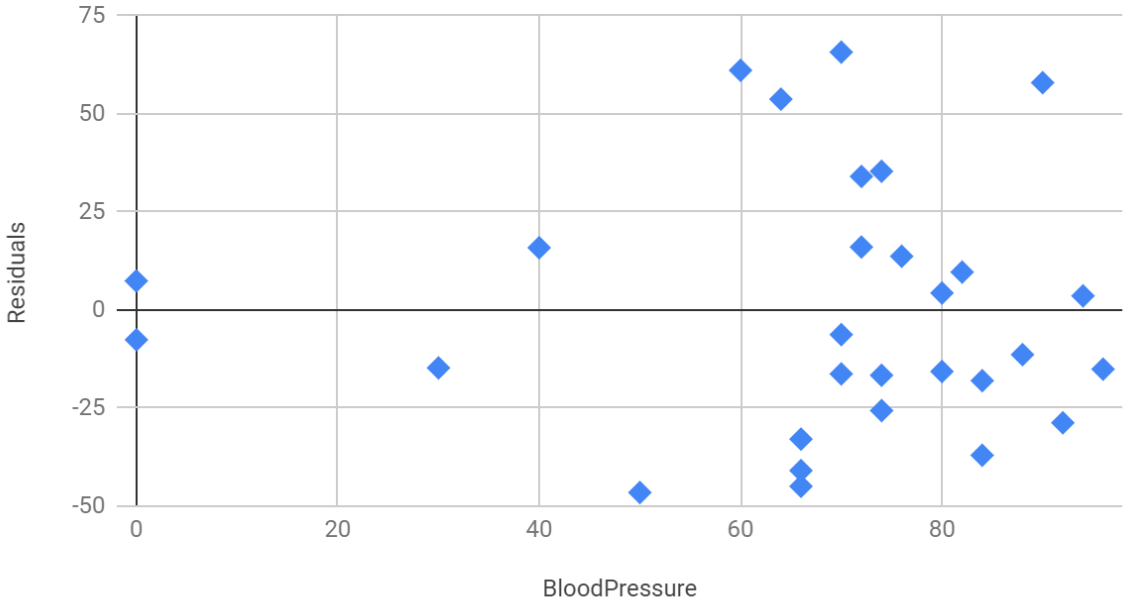
RESIDUAL OUTPUT

Observation	*Predicted Glucose*	*Residuals*
1	131.9648238	16.03517623
2	129.9335388	-44.9335388
3	129.2564438	53.74355619
4	129.9335388	-40.9335388
5	121.1313039	15.86869609
6	132.6419188	-16.64191877
7	124.5167789	-46.51677887
8	107.5894041	7.410595912
9	131.2877288	65.71227122
10	140.0899637	-15.08996367
11	138.7357737	-28.73577369
12	132.6419188	35.35808123
13	134.6732037	4.32679626
14	127.9022538	61.09774617
15	131.9648238	34.03517623
16	107.5894041	-7.589404088
17	136.0273937	-18.02739372
18	132.6419188	-25.64191877
19	117.745829	-14.74582896
20	131.2877288	-16.28772878
21	137.3815837	-11.3815837
22	136.0273937	-37.02739372
23	138.0586787	57.9413213
24	134.6732037	-15.67320374
25	139.4128687	3.587131321
26	131.2877288	-6.287728783
27	133.3190138	13.68098624
28	129.9335388	-32.9335388
29	135.3502987	9.649701269

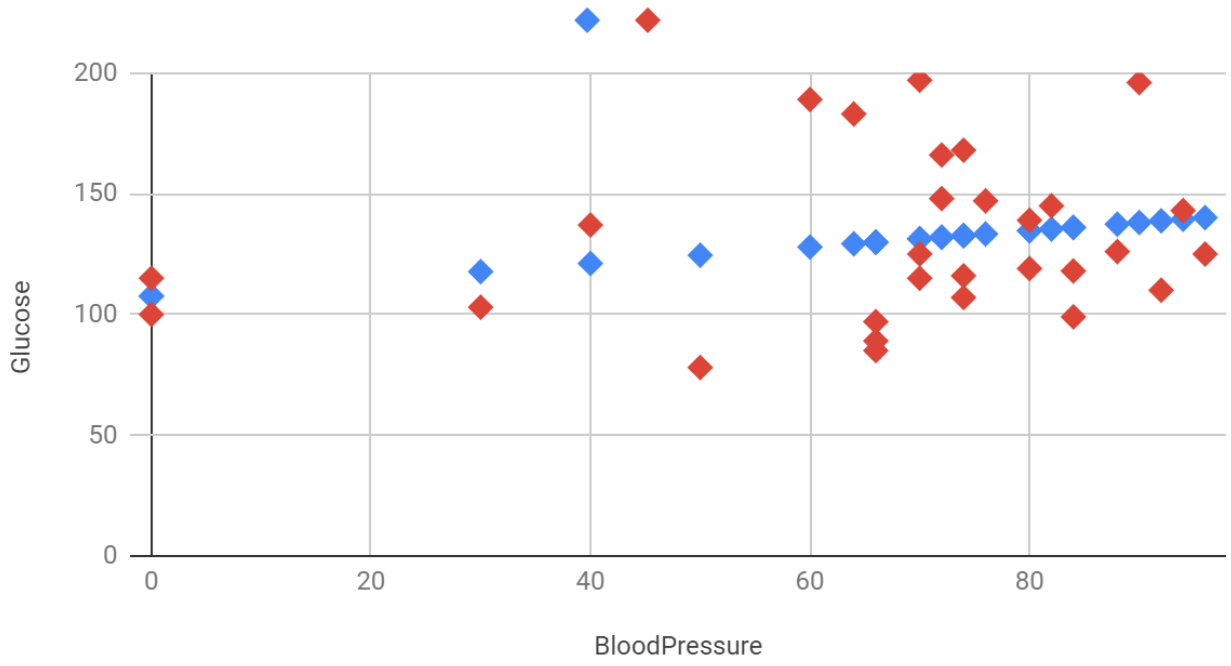
PROBABILITY OUTPUT

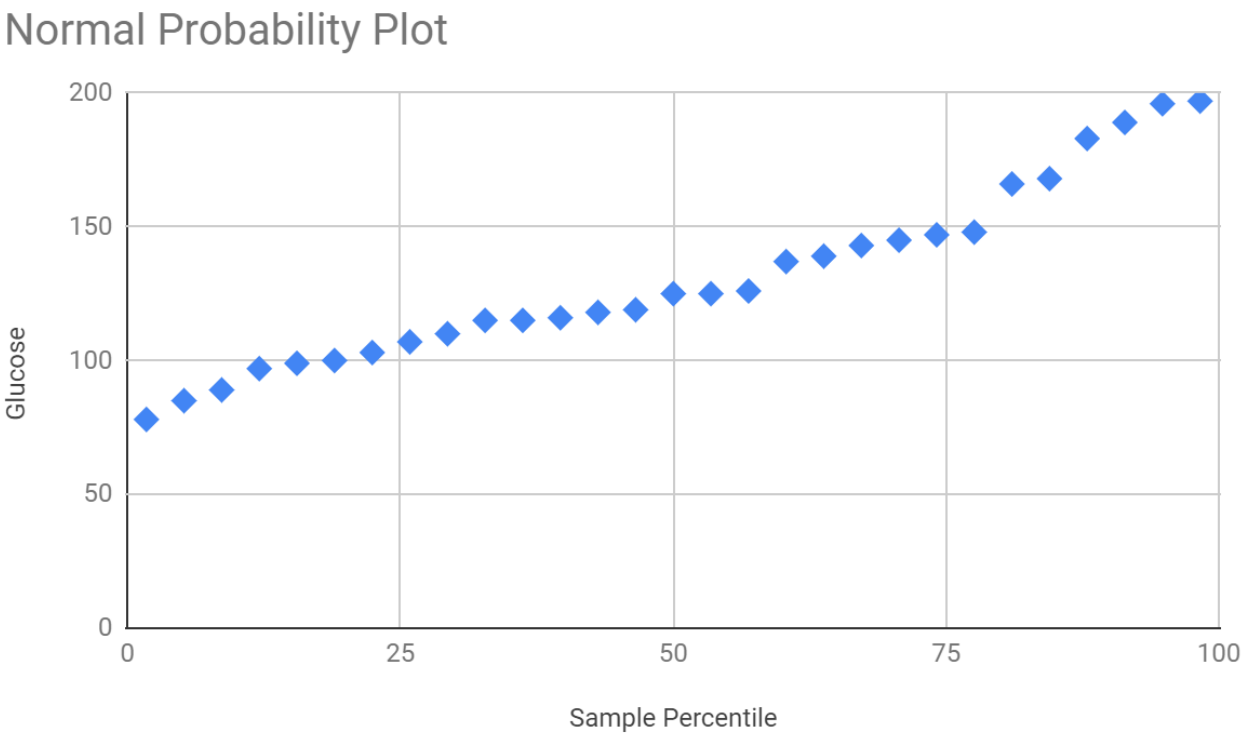
Percentile	*Glucose*
1.724137931	78
5.172413793	85
8.620689655	89
12.06896552	97
15.51724138	99
18.96551724	100
22.4137931	103
25.86206897	107
29.31034483	110
32.75862069	115
36.20689655	115
39.65517241	116
43.10344828	118
46.55172414	119
50	125
53.44827586	125
56.89655172	126
60.34482759	137
63.79310345	139
67.24137931	143
70.68965517	145
74.13793103	147
77.5862069	148
81.03448276	166
84.48275862	168
87.93103448	183
91.37931034	189
94.82758621	196
98.27586207	197

BloodPressure Residual Plot



BloodPressure Line Fit Plot




Normal Probability Plot
200
150
100
50
0
Glucose
0
25
50
75
100
Sample Percentile