॥वसुधैव कुटुम्बकम्॥

SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE

Symbiosis International (Deemed University)

(Established under section 3 of the UGC Act, 1956)

Re-accredited by NAAC with 'A' grade (3.58/4) | Awarded Category – I by UGC

Founder: Prof. Dr. S. B. Mujumdar, M. Sc., Ph. D. (Awarded Padma Bhushan and Padma Shri by President of India)

Assignment No. 12

Subject:	Software Tools for Computer Science Engineering (STCS)
Name of Student	**Almaan Hadwani**
PRN No.	**24070122021**
Branch	CSE (A1)
Academic Year & Semester	2024-2025 & Semester 1
Date of Performance	**04-12-2024**
Title of Assignment:	To study Regression and its types by using MS-Excel

Theory

Write a note on Regression analysis.

Regression analysis is a statistical technique used to examine the relationship between one or more independent variables (predictors) and a dependent variable (outcome). Its primary purpose is to model and analyze the relationship between variables to predict the dependent variable, understand underlying patterns, or assess the strength of relationships. Regression analysis is a fundamental tool in statistics and data science, providing insights into relationships between variables, helping to make predictions, and guiding decision-making in various fields. Understanding its assumptions, strengths, and limitations is crucial for building reliable models.

Mention different types of regression.

1. Linear Regression:

Simple Linear Regression: Involves one independent variable and one dependent variable, assuming a linear relationship.

2. Logistic Regression:

Used when the dependent variable is binary (0 or 1). It models the probability of an event occurring using the logistic function, not a linear relationship.

3.Polynomial regression:

It is an extension of linear regression that allows for modeling non-linear relationships between the

independent polynomial. regression fits a curve to the data by including polynomial terms (higher powers of the independent variable) in the model.

4.Ridge regression:
It is a type of regularized linear regression that helps address the issue of multicollinearity (high correlation between predictor variables) and overfitting, which can occur when the model is too complex or the number of predictors is very large. It achieves this by adding a penalty term to the ordinary least squares (OLS) loss function, which helps shrink the regression coefficients, reducing their magnitude and making the model more robust.

5.Lasso regression (Least Absolute Shrinkage and Selection Operator) :
It is a regularized version of linear regression that, like ridge regression, is used to prevent overfitting and handle multicollinearity. However, unlike ridge regression, lasso uses L1 regularization instead of L2 regularization. This difference has a significant impact on the behavior of the model, particularly in terms of feature selection.

6.Quantile regression is a type of regression analysis used to estimate the conditional quantiles (such as the median, or other percentiles) of a response variable, rather than the conditional mean, as in ordinary least squares (OLS) regression.

7.Bayesian linear regression is a type of conditional modeling in which the mean of one variable is described by a linear combination of other variables, with the goal of obtaining the posterior probability of the regression coefficients.

8.Principal component regression (PCR) is a regression analysis technique that is based on principal component analysis (PCA). PCR is a form of reduced rank regression. More specifically, PCR is used for estimating the unknown regression coefficients in a standard linear regression model

Explain Linear Regression in detail.
Linear regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. The primary objective of linear regression is to find the line (or hyperplane in the case of multiple variables) that best fits the data, thereby allowing predictions and insights into relationships between variables.
Linear regression assumes that the relationship between the dependent variable it can be described by a straight line in the simplest case of one predictor, or a plane/hyperplane in the case of multiple predictors.

Create 2 vectors x and y. Calculate the equation for linear regression. Plot the scatter graph for the given data and the calculated data.

Output

Paste the screenshots for all mentioned above.

Book2 - Excel

SUMMARY OUTPUT								
Regression Statistics								
Multiple F	0.987497							
R Square	0.97515							
Adjusted I	0.974622							
Standard I	487.6155							
Observati	49							

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	4.39E+08	4.39E+08	1844.372	2.28E-39
Residual	47	11175137	237768.9		
Total	48	4.5E+08			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	218.7627	132.1006	1.656031	0.104378	-46.9894	484.5149	-46.9894	484.5149
X Variable	202.0502	4.704732	42.94615	2.28E-39	192.5854	211.5149	192.5854	211.5149

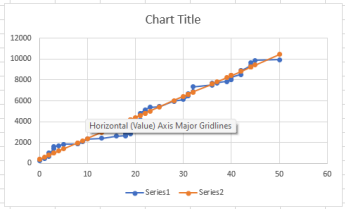
m	
c	
e	

y=194.11*x-26.426*26.149

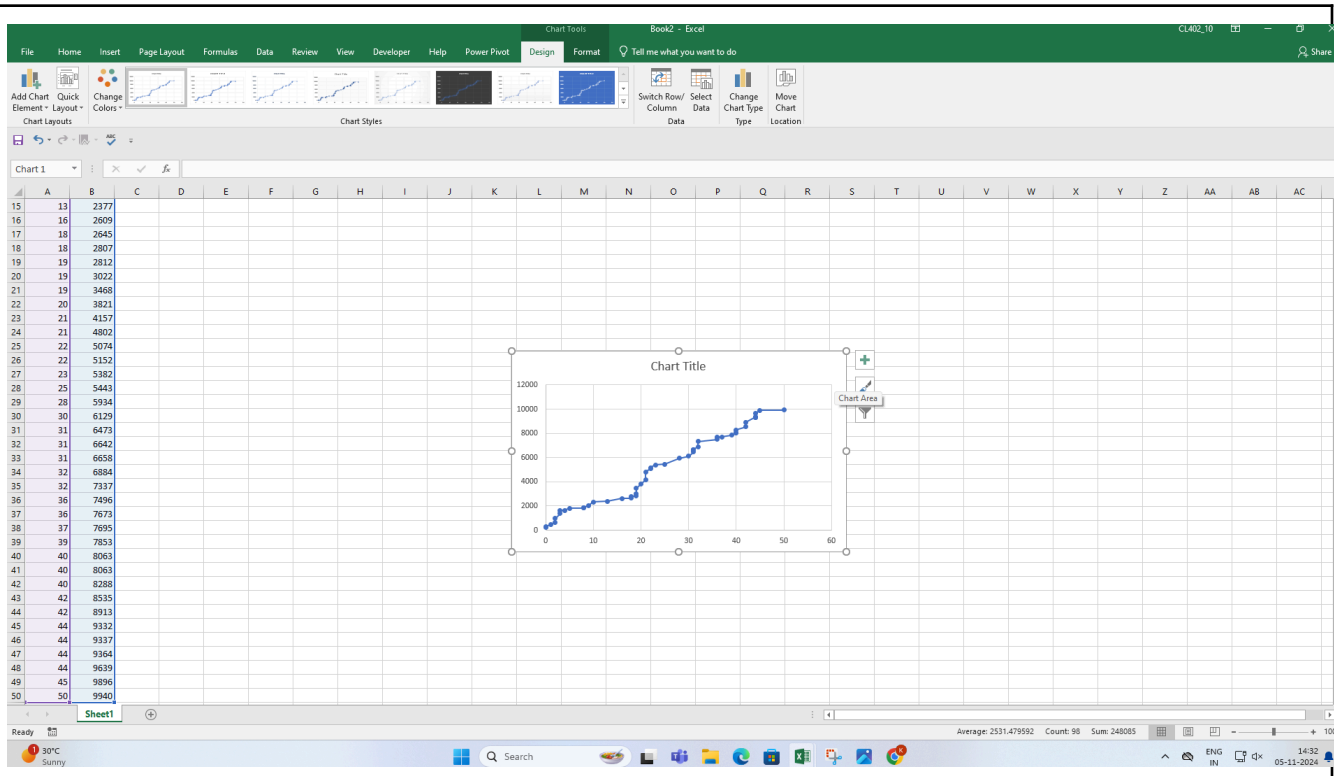
Sheet2 Sheet1

Book2 - Excel

13	2377	2977.516
16	2609	3583.667
18	2645	3987.767
18	2807	3987.767
19	2812	4189.817
19	3022	4189.817
19	3468	4189.817
20	3821	4391.867
21	4157	4593.918
21	4802	4593.918
22	5074	4795.968
22	5152	4795.968
23	5382	4998.018
25	5443	5402.118
28	5934	6008.269
30	6129	6412.369
31	6473	6614.42
31	6642	6614.42
31	6658	6614.42
32	6884	6816.47
32	7337	6816.47
36	7496	7624.671
36	7673	7624.671
37	7695	7826.721
39	7853	8230.821
40	8063	8432.871
40	8063	8432.871
40	8288	8432.871
42	8535	8836.972
42	8913	8836.972
44	9332	9241.072
44	9337	9241.072
44	9364	9241.072
44	9639	9241.072
45	9896	9443.122
50	9940	10453.37



Sheet2 Sheet1



Conclusion

Regression analysis in MS Excel simplifies understanding data relationships. It supports various regression types, predictive modeling and analytical skills effectively.

Faculty Name: Prof. Shubhangi Deokar