Business Analytics

BSMS2002

TA sessions
Utkarsh Sahu (21f1001107@ds.study.iitm.ac.in)

https://github.com/utkarsh4tech/BSMS2002

Tag us on discourse

1. Prof. Dr. Malolan Sundararaman - @Milo
2. Prof. Swaminathan - @ram158
3. TA Utkarsh Sahu - @SahuUtkarsh03

In Multiple Linear Regression, the "R" represents _____ (choose all those that are applicable)

SLR → Cor(x, y)

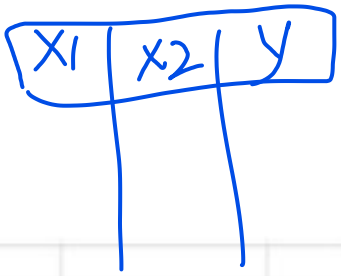
a. Correlation between the dependent variable and all independent variables

b. Correlation between the actual and predicted values of the dependent variable

c. Correlation between the predicted value of the dependent variable and the actual value of the independent variable

d. Correlation between the errors

Find A1, A3 and A4



$Y = B_0 + B_1 X_1 + B_2 Y_2$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	1.92403E-15
Observations	A1

ANOVA

	df	SS	MS	F	Significance F
Regression	A3	905.7333333	A4		
Residual	12				
Total	14				

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	7.10543E-15	3.48311E-15			-4.83607E-16	1.46945E-14	-4.83607E-16	1.46945E-14
X Variable 1	2	1.57612E-16			2	2	2	2
X Variable 2	3	2.25138E-16			3	3	3	3

1) How many columns are present in the sample used to develop the regression model (all the columns present in the sample)?

3

2) Which of the following variables are “NOT Significant”? (choose all that is applicable)

a. X1

b. X2

c. None of the above

A multiple linear regression model, as specified below, is fit on a dataset with 250 data points. Then answer the given subquestions

$$\text{MLR Model: } Y = 2.1 + 1.4 * X_1 - 4.2 * X_2 + 0.5 * X_3 + 7 * X_4 + \varepsilon$$

1) How many degrees of freedom are present for the "Residuals" in the ANOVA Table? 245 (=249−4)
2) How many total degrees of freedom are present for the fitted model in the ANOVA Table? 249
3) If no feature engineering was performed, then how many features were present in the dataset? 4

You are solving a regression problem with 8 explanatory variables. The data has 150 observations, and the R-square value was found to be 0.75. You are adding one more explanatory variable to the dataset (a total of 9 explanatory variables). The new R-square value is 0.8, and the new adjusted R-square value is 0.92. What does this imply?

a. The new variable does not improve the model

b. The new variable improve the model

c. None of the above

Old Adj $R^2 = 0.735$

New " " $= 0.78$

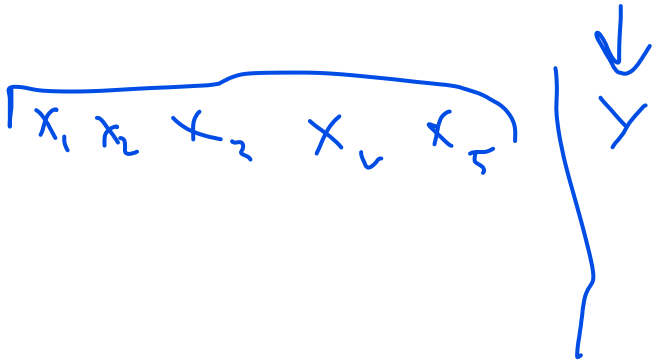
What does the term "Multicollinearity" refer to? (Select all that are applicable)

a. The dependent and independent variables are not-related

b. The dependent and independent variables are linearly related

c. The dependent variable is linearly related to another dependent variable

d. None of these



$1 - [(1 - R^2)(n-1)/(n-k-1)]$ k

You are solving a regression problem with 4 explanatory variables. The data has 40 observations and the R-square value was found to be 0.74.

n

1) What is the value of adjusted R-square 0.71 ✓

2) You are adding a new explanatory variable to the dataset and the new adjusted R squared value is 0.745. Is the new variable significant? ✓ Yes

3) You are removing a few explanatory variables from the dataset and the new adjusted R square value is -0.21. Is it possible?

You are conducting a multiple linear regression with sales as the dependent variable. Price, quantity and rating score are the independent variables. In order to calculate the VIF for the variable rating score, you implement a linear regression with rating score as the dependent variable and other variables as independent variables and obtain R-squared of 0.3. What is the VIF for the variable rating score?

$$\frac{1}{0.7} = ?$$

In Regression, “Marginal” slope and “Partial” slope coincide if the explanatory variables are dependent.

a. True

b. False

You are conducting a multiple linear regression with sales as the dependent variable. Price, quantity and rating score are the independent variables. In order to calculate the VIF for the variable rating score, you implement a linear regression with rating score as the dependent variable and other variables as independent variables and obtain R-squared of 0.3. What is the VIF for the variable rating score?

Milo’s Motors (MM) is a motorcycle brand that manufactures electric two wheelers. MM wants to understand the relationship between “Mileage”, “Storage space” and “Charging Time in Minutes” on “Sales volume”. The owner of the company, Dr. Milo, only has half-baked knowledge of regression. Hence, several regression models (Model-1, Model-2, Model-3, Model-4, Model-5, Model-6, Model-7, Model-8, Model-9 and Model-10) which are given below were built on appropriate available data by Dr. Milo.Given this information, answer the subquestions.

Click Here for Models

1) How many observations (rows) are present in the data set used to build Model-1

2) What is the total indirect effect of "Mileage" on "Sales Volume"?

3) What is the value of the "T-statistic" associated with the intercept in Model-3?

4) What is the adjusted R-Square value for Model-10?

5) What is F-statistic for Model-10?

6) For which of the following "Tabulated F" values, will the Null Hypothesis NOT BE REJECTED for Model-10?(Choose all that is applicable)

a. 240 b. 350 c. 480 d. 520