

**LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034****M.Sc. DEGREE EXAMINATION – STATISTICS****FIRST SEMESTER – NOVEMBER 2022****PST1MC02 – APPLIED REGRESSION ANALYSIS**

Date: 25-11-2022

Dept. No. 

Max. : 100 Marks

Time: 01:00 PM - 04:00 PM

**SECTION - A****Answer ALL the questions**

<b>1</b>	<b>Fill in the blanks / Answer the following / MCQ/Definition</b>	<b>(5 x 1 = 5)</b>	
a)	If we add a feature in linear regression model and retrain the same model. Which of the following option is true R square? a) If R Squared increases, this variable is significant b) If R Squared decreases, this variable is not significant c) Individually R squared cannot tell about variable importance. We can't say anything about it right now d) None of these	K1	CO1
b)	Multicollinearity refers to a situation in which a) Successive error terms derived from the application of regression analysis to time series data are correlated. b) There is a high degree of correlation between the independent variables included in a multiple regression model. c) The dependent variable is highly correlated with the independent variable(s) in a regression analysis. d) The application of a multiple regression model yields estimates that are nonlinear in form.	K1	CO1
c)	<b>In regression modelling, the impacts of trade-off between under-fitting and over-fitting the most can be identified through.</b> a) The polynomial degree b) The weights by matrix inversion or gradient descent c) The use of a constant-term d) None of these	K1	CO1
d)	In a regression analysis if $SSE = 200$ and $SSR = 300$ , then the coefficient of determination is a) 0.6667      b) 0.6000      c) 0.4000      d) 1.5000	K1	CO1
e)	The Durbin-Watson statistic is used to test for-----	K1	CO1

2	<b>Match the following / True or False/ Fill in the blanks / Answer the following / MCQ/Definitions</b> <b>(5 x 1 = 5)</b>		
a)	The $R^2$ of a model with 3 regressors built with 25 observations is 0.65, Summaize Adjusted $R^2$ of the model.	K2	CO1
b)	illustrate the variance-stabilizing transformation for a Poisson count variable	K2	CO1
c)	Explain any two sources of multicollinearity.	K2	CO1
d)	Classify the term hierarchical model.	K2	CO1
e)	Demonstrate a 'Random Walk Without Drift'.	K2	CO1
<b>SECTION - B</b>			
<b>Answer any THREE of the following questions.</b>			<b>(3 x 10 = 30)</b>
3	Construct five different scenarios that can show up in plotting residuals versus the fitted values and Identify how these plots help in detecting model inadequacies.	K3	CO2
4	Utilize Generalized Least Squares and develop the estimation of the regression parameters and ANOVA. Discuss Weighted Least Square and the issues related to using WLS.	K3	CO2
5	In a regression model-building study, the subjects were classified into four categories according to mode of commute to office (bus/ train/ two-wheeler/four-wheeler) and there was a single numerical variable 'distance of travel'. The analyst wishes to allow the possibility of different intercepts and slopes for the four classes. List out the columns (variables) of the data matrix. Organize the explicit equations for the four classes.	K3	CO2
6	Develop Non-parametric regression through 'Kernel Smoothing' and list out any two kernel functions.	K3	CO2
7	Define autocorrelation and Identify the reasons for autocorrelation.	K3	CO2
<b>SECTION - C</b>			
<b>Answer any TWO of the following questions.</b>			<b>(2 x 12.5 = 25)</b>
8	A linear regression model with an intercept term and 4 independent variables was built using 100 observations. It was reported that $\sum Y_i = 540$ , $\sum Y_i^2 = 8100$ , $\sum \hat{Y}_i^2 = 6750$ . Construct the ANOVA table and carry out the test for the overall significance of the model.	K4	CO3
9	Classify the AR(p), MA(q), ARMA(p,q), I(d), ARIMA(p,d,q) processes.	K4	CO3
10	List the methods of detecting outliers in detail.	K4	CO3
11	Take part in the Box-Cox class of power transformations and inspect the analytical method of choosing the power.	K4	CO3
<b>SECTION - D</b>			
<b>Answer any ONE of the following questions.</b>			<b>(1 x 15 = 15)</b>
12	Determine different subset models with a sample of size 20 by applying the 'Forward	K5	CO4

method to build a model with four regressors given the following information on  $SS_{Res}$  for Use a significance of 5%:  
 $SS_{Total} = 4752.58$ ,  $SS_{Res}(X_1) = 1546.79$ ,  $SS_{Res}(X_2) = 2214.97$ ,  $SS_{Res}(X_3) = 1586.06$ ,  
 $SS_{Res}(X_4) = 3393.95$ ,  $SS_{Res}(X_1, X_2) = 130.83$ ,  $SS_{Res}(X_1, X_3) = 1520.54$ ,  $SS_{Res}(X_1, X_4) = 307.55$ ,  
 $SS_{Res}(X_2, X_3) = 101.36$ ,  $SS_{Res}(X_2, X_4) = 2147.36$ ,  $SS_{Res}(X_3, X_4) = 727.02$ ,  $SS_{Res}(X_1, X_2, X_3) = 83.97$ ,  
 $SS_{Res}(X_1, X_2, X_4) = 88.97$ ,  $SS_{Res}(X_1, X_3, X_4) = 129.15$ ,  $SS_{Res}(X_2, X_3, X_4) = 84.21$ ,  
 $SS_{Res}(X_1, X_2, X_3, X_4) = 83.76$

13 Choose the Durbin-Watson test to determine whether first-order autocorrelation exists [from the following information](#) of OLS residuals:  
 0.18   -0.21   1.25   2.10   1.55   -2.05   0.8   -0.64   -1.46   0.11   -0.85   0.10  
 -1.44   0.58   -0.08. [It is given that  $d_L = 1.08$ ,  $d_U = 1.36$ ]

K5   CO4

**SECTION - E**

**Answer any ONE of the following questions. (1 x 20 = 20)**

14 Build the regression model  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$  for the data given below and complete the ANOVA table. Also comment on your findings.

Score in the final (Y)	First Preliminary (X1)	Second Preliminary (X2)
68	78	73
85	82	79
90	90	92
55	68	67
65	70	66
75	79	75
81	89	84
79	81	82

K6   CO5

15 Develop the various methods of diagnosing multicollinearity and suggest the methods for removing it.

K6   CO5

@@@@@