LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **STATISTICS**

FIRST SEMESTER – **NOVEMBER 2023**

PST1MC02 – APPLIED REGRESSION ANALYSIS

Date: 03-11-2023 Dept. No. Max. : 100 Marks

Time: 01:00 PM - 04:00 PM

	SECTION A – K1 (CO1)	
	Answer ALL the questions	**(5 x 1 = 5)**
1	**MCQ**	
a)	Consider two regression models (1 and 2) with R^2 of 0.52 and 0.89 respectively. Which among the following statements is true? a) Goodness of fit in regression 1 is more than that of 2 b) Goodness of fit in regression 2 is more than that of 1 c) Both (a) and (b) are true d) None of the above	
b)	WLS is used to correct for: a) Multicollinearity b) Autocorrelation c) Hetroscedasticity d) Correlation	
c)	Variance Inflation Factor is used for a) Detecting Hetroscedasticity b) Solving Hetroscedasticity c) Detecting Multi-collinearity d) Solving Multi-collinearity	
d)	If the value of Durbin-Watson's d = 0, there is a) No Auto-correlation b) Positive Auto-correlation c) Negative Auto-correlation d) None of these	
e)	Regression models containing a mixture of quantitative and qualitative variables are called: a) ANOVA models. b) ANCOVA models. c) Parallel regressions. d) Coincident regressions.	
	SECTION A – K2 (CO1)	
	Answer ALL the questions	**(5 x 1 = 5)**
2	**True or False/Fill in the blanks**	
a)	Say True or False: The statistical properties of OLS estimators are Linearity, Unbiasedness and minimum variance	
b)	The conditional mean of Y is-------	
c)	Assumption of 'No multicollinearity' means the correlation between the regressand and regressor is----------	
d)	Say True of Fall\se: Even if heteroscedasticity is suspected and detected, it is not easy to correct the problem.	
e)	In regression model, one of the explanatory variables included is the lagged value of the dependent variable, then the model is referred to as --------	
	SECTION B – K3 (CO2)	
	Answer any THREE of the following	**(3 x 10 = 30)**
3	A model with five records was built where the Y values were 1.7, 3.5, 2.9, 3.1, 2.5 and $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & -1 & 3 & 2 & -2 \end{bmatrix}$. Construct the vector of residuals by make use the 'Hat' matrix.	
4	A linear regression model with an intercept term and 4 independent variables was built using 100 observations. It was reported that $\sum Y_i = 540$, $\Sigma Y_i^2 = 8100$, $\Sigma \hat{Y}_i^2 = 6750$. Construct the ANOVA table and carry out the test for the overall significance of the model.	

5	Develop and organize the orthogonal polynomial fitting of regression model in detail.
6	Explain the various methods of diagnosing multicollinearity
7	Classify the AR(p), MA(q) and ARMA(p, q) processes.

SECTION C – K4 (CO3)

	Answer any TWO of the following	**(2 x 12.5 = 25)**
8	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.	
9	Classify and compare the following model building procedures. a. Forward Selection b. Backward Elimination c. Stepwise procedure	
10	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.	
11	The residuals from a model arranged in time-order had the following signs + + + + – – – – – – – + – – + + – – – + + + – – + + – – – +. Apply the 'Runs Test' in the context of time-series residual analysis.	

SECTION D – K5 (CO4)

	Answer any ONE of the following	**(1 x 15 = 15)**
12	a) Take part in the Box-Cox class of power transformations and inspect the analytical method of choosing the power. (8) b) Design General Linear Hypothesis and develop the F-test for it. Discuss the test for the linear hypothesis H_0: $\beta_2 = \beta_3$ for a linear model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$,. (7)	
13	a) Compose the methods of studying autocorrelation in a regression model. (8) b) Elaborate the Box-Jenkins methodology of ARIMA modelling. (7)	

SECTION E – K6 (CO5)

	Answer any ONE of the following	**(1 x 20 = 20)**
14	a) The ANOVA table for testing overall significance of the model coefficients is given below. Invert the missing entries	

Source of Variation	Sum of Squares	Degrees of Freedom	Mean sum of Square	F - Ratio
Regression	?	3	47	?
Error	1643	?	?	-
Total	?	139	-	-

Also Estimate the R^2 and adjusted R^2 using the above ANOVA table (6)

b) Criticize the Durbin-Watson Statistics to test for first order autocorrelation in the time ordered error terms obtained by OLS with three regressors of the model.
4.818, -10.364, 4.454, -0.727, 4.091, -1.092, -6.272, 3.546, 8.364, -6.818
The relevant DW bound are given to be $dL = 0.34$, $dU = 1.733$ (14)

15 **a)** Determine different subset models with a sample of size 20 by applying the 'Forward method to build a model with four regressors given the following information on SS_{Res} for Use a significance of 5%: (15)
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

b) 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. (6)