



**LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034**

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

FIRST SEMESTER – NOVEMBER 2016

**16PST1MC02 / ST 1821 - APPLIED REGRESSION ANALYSIS**

Date: 04-11-2016  
Time: 01:00-04:00

Dept. No.

Max. : 100 Marks

**SECTION – A: ANSWER ALL QUESTIONS**

**(10 X 2 = 20 )**

- 1 Find the regression coefficient and also write down the multiple regression from the following  

$$(X'X)^{-1} = \begin{bmatrix} 25 & 219 & 10,232 \\ 219 & 3,055 & 133,899 \\ 10,232 & 133,899 & 6,725,688 \end{bmatrix}$$
 and  $X'Y = \begin{bmatrix} 559.60 \\ 1,375.44 \\ 337,072.00 \end{bmatrix}$
- 2 Define PRESS statistics and R<sup>2</sup> prediction based on PRESS.
- 3 Write the linear transformation for  $y = x/(60x - 10)$
- 4 What do you mean by indicator variable?
- 5 Mention any two criteria for evaluating subset in regression model building.
- 6 Find out condition number and condition indices of the matrix from the following:  
 $\lambda_1 = 4.2048, \lambda_2 = 2.1626, \lambda_3 = 41.1384, \lambda_4 = 0.0001$
- 7 What do you mean by hierarchical regression model?
- 8 Write down least square normal equations for nonlinear regression model  $y = \theta_1 + \theta_2x + \theta_3x^2 + \epsilon$
- 9 Explain the term auto correlation.
- 10 Explain AR (1) process.

**SECTION – B: ANSWER ANY FIVE QUESTIONS**

**(5 X 8 = 40 )**

- 11 A model with four predictors was built where the Y values were 1.7, 3.5, 2.9, 3.1, 2.5 and the data matrix was  $X = \begin{bmatrix} 4 & 1 & 11 & 1 \\ 1 & 1 & 11 & 1 \end{bmatrix}$ . Compute the vector of residuals by computing the 'Hat' Matrix.
- 12 In multiple regression modelling, give rough graphical illustrations of the different possible scenarios when the residuals are plotted against the predicted values. Describe how these are used for model modification
- 13 Explain the Box-Cox class of power transformations and describe the practical method of choosing the power.
- 14 Following are the part of output obtained in the singular values analysis and variance decomposition proportions were carried out to detect Multicollinearity.

Eigen values	Singular values	Condition indices	Variance decomposition proportions				
			intercept	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
2.63287	1.622612	-	-	0.0568	0.0329	0.0036	0.0049
1.03335	-	1.96214	0.0001	0.1473	0.0696	0.1159	-
-	0.778287	2.084852	0.0032	0.6325	0.0869	-	0.0027
-	-	2.300062	0.0001	-	0.1074	0.2196	0.2105
0.00093	0.00496	-	0.9964	0.0588	-	0.0205	0.6645

Fill up the missing entries and identify the variables entangled in collinear relationship.

- 15 What are the points to be considered in fitting a polynomial regression model?
- 16 Define the term interaction effect and illustrate with an example. How the interaction effect of two categorical variables is captured by the coefficient of product term?
- 17 Describe 'Unit Root Test' for stationarity of a time-series.
- 18 Give a brief note about 'Ridge Regression' with example.

**SECTION – C: ANSWER ANY TWO QUESTIONS**

**(2 X 20 = 40 )**

- 19 a) Following are the multiple linear regression model output using excel:

Output 1: ANOVA

	df	SS	MS	F	Significance F
Regression	2	21.60055651	10.80028	32.87837	0.000276
Residual	7	2.299443486	0.328492		
Total	9	23.9			

Output 2:

	Coefficients	Stand. Error	t Stat	P-value	Lower	Upper
Intercept	-0.8687014	0.951547725	-0.9129	0.39163	-3.1187	1.3813
Miles Travelled	0.06113459	0.009888495	6.18239	0.00045	0.03775	0.0845
No. of Deliveries	0.92342536	0.221113461	4.17625	0.00415	0.40057	1.4462

Output 3:

S = 0.57314                      R-square = 0.903788975      Adj, R-square = 0.87630011

Write down the multiple regression model and Provide interpretation for (i) Model coefficients (ii) Overall significance of the model fit (iii) Significance of the individual regression coefficient and (iv) interpret R-square (Travel time in hours is response variable)

- 19 b) Explain Studentized Residuals and externally studentized Residuals.
- 20 a) Explain Non-parametric regression through 'Kernel Smoothing' and list out any two kernel functions.
- 20 b) Explain spline regression with an example.
- 21 a) What are the various methods of diagnosing multicollinearity and recommend the methods for removing it
- 21 b) Given the following information for fitting a regression model with 4 regressors. Use forward selection method to build a regression model.

$SS_T = 2715.7635$	$SS_{Res}(x_1, x_3) = 1227$	$SS_{Res}(x_1, x_2, x_4) = 47.9$
$SS_{Res}(x_1) = 1265.6867$	$SS_{Res}(x_1, x_4) = 74.76$	$SS_{Res}(x_1, x_3, x_4) = 50.8$
$SS_{Res}(x_2) = 906.3363$	$SS_{Res}(x_2, x_3) = 415.4$	$SS_{Res}(x_2, x_3, x_4) = 73.8$
$SS_{Res}(x_3) = 1939.4$	$SS_{Res}(x_2, x_4) = 868.8$	$SS_{Res}(x_1, x_2, x_3, x_4) = 47.86$
$SS_{Res}(x_4) = 883.87$	$SS_{Res}(x_3, x_4) = 175.7$	
$SS_{Res}(x_1, x_2) = 57.9$	$SS_{Res}(x_1, x_2, x_3) = 48.1$	

- 22 a) Define the Durbin-Watson statistics to test for first order autocorrelation in the error terms of a model. Apply it to the following series of time ordered residuals obtained by least square for a model with three regressors:

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  $d_L = 0.34$  and  $d_U = 1.733$

- 22 b) Explain the Box-Jenkins methodology of ARIMA modelling.

\*\*\*\*\*