LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

FIRST SEMESTER – **APRIL 2016**

ST 1821 - APPLIED REGRESSION ANALYSIS

Date: 30-04-2016 Dept. No. Max. : 100 Marks
Time: 01:00-04:00

SECTION – A : ANSWER ALL QUESTIONS (10 X 2 = 20)

What do you mean by standardized regression coefficients?

Write down the multiple linear regression equation and the standardized regression coefficients based on the following information $(X'X)^{-1} = \begin{bmatrix} 4 & 12 & -4 \\ 12 & 14 & -8 \\ -4 & -8 & 6 \end{bmatrix}, XY = \begin{bmatrix} 6 \\ 14 \\ 20 \end{bmatrix}$

What is the use of adjusted R-square?

Identify the linearizing transformation required to transform the relation $Y = X/(\beta_0 X - \beta_1)$ and write down the linearized form.

State the variance-stabilizing transformation for a Poisson count variable.

What is a hierarchical model?

Write down the equations satisfied by the residuals that are produced by building a regression model for a response variable Y with a single regressor X.

Define General Linear Model (GLM).

Briefly explain any one source for auto correlation in time series data.

What are the steps involved in Box-Jenkins methods?

SECTION – B: ANSWER ANY FIVE QUESTIONS (5X 8 = 40)

In multiple regression modeling, 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.

Explain general linear hypothesis and develop the F-test for it. For a linear model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$ develop the test for the linear hypothesis $H_0: \beta_2 = \beta_3$.

Explain the Box-Cox class of power transformations and state the appropriate forms relevant for model comparison. Describe the practical method of choosing the power.

Define the term interaction effect and illustrate with an example how the interaction effect between two categorical explanatory variables is captured by the coefficient of the cross product term?

Explain Non-parametric regression using 'Kernal Smoothing' mentioning the different choices for the Kernal functions.

Explain four sources of multicollinearity with examples.

Give a brief note about 'Ridge Regression' with example.

Describe Unit Root Test for stationarity of a time series.

SECTION – C: ANSWER ANY TWO QUESTIONS (2X 20= 40)

19 The weight and systolic blood pressure of randomly selected males in the age group of 25-30 are given below: (20)

Weight	162	164	186	154	222	180	193	214	204

[illegible]	133	132	145	133	156	166	140	134	149

Find a regression line relating systolic blood pressure to weight. Test the hypothesis $H_0: \beta_1 = 0$. Also calculate R^2 and 95% confidence interval on the slope.

20 (a) In building a model with four regressors, the singular values analysis and variance decomposition proportions were carried out to detect multicollinearity and the part of output obtained in the analysis below.

Eigen values (of X'X)	Singular values (of X)	Condition Indices	Variance decomposition proportions				
			Intercept	X_1	X_2	X_3	X_4
2.63287	1.622612			0.0568	0.0329	0.0036	0.0049
1.03335		1.596214	0.0001	0.1473	0.0696	01159	
	0.778287	2.084852	0.0032	0.6325	0.0869		0.0027
		2.300062	0.0001		0.1074	0.2196	0.2105
0.00093	0.030496		0.9964	0.0588		0.0205	0.6645

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

b) Define the Durbin-Watson Statistic 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 OLS for a model with three regressors:

(10)

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 $d_L = 0.34$, $d_U = 1.733$

(10)

21 What do you mean by an indicator variable? Also explain about their advantages and illustrate different intercepts and different slopes for three classes. (20)

22 (a) Carryout the forward model building process to build a model with four regressors given the following information on SS_{Res} for different subset models with a sample of size 20. Use a significance of 5%.

$SS_{total} = 271.57$

$SS_{Res}(X_1) = 126.569$ $SS_{Res}(X_1, X_2) = 122.707$ $SS_{Res}(X_1, X_2, X_3) = 4.797$

$SS_{Res}(X_2) = 90.633$ $SS_{Res}(X_1, X_3) = 7.476$ $SS_{Res}(X_1, X_2, X_4) = 5.084$

$SS_{Res}(X_3) = 193.94$ $SS_{Res}(X_1, X_4) = 44.544$ $SS_{Res}(X_1, X_3, X_4) = 7.381$

$SS_{Res}(X_4) = 88.387$ $SS_{Res}(X_2, X_3) = 86.888$ $SS_{Res}(X_2, X_3, X_4) = 4.786$

$SS_{Res}(X_1, X_2) = 5.791$ $SS_{Res}(X_2, X_4) = 17.574$

$SS_{Res}(X_1, X_2, X_3) = 4.811$

(15)

b) Explain how the estimates of regression parameters and the joint and individual significance of the regressors are useful to detect the presence of collinearity. (5)