LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc.- DEGREE EXAMINATION- **STATISTICS**

FIRST SEMESTER- **NOVEMBER 2015**

ST 1821 – APPLIED REGRESSION ANALYSIS

Date: 05/11/2015
Time: 01:00-04:00 DEPT. NO. MAX. : 100 MARKS

ANSWER **ALL** THE QUESTIONS SECTION A (10 X 2 = 20)

1 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.

2 State any two applications of Multiple Linear Regression model.

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

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

5 Write a note on the 'Model Respecification' method of handling multicollinearity.

6 Write down the multiple linear model equation and intercept of the model coefficients based on the following information

$$(X'X)^{-1} = \begin{bmatrix} 2 & 6 & -2 \\ 6 & 7 & -4 \\ -2 & -4 & 3 \end{bmatrix}, X'Y = \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}$$

7 What is a hierarchical model?

8 Define General Linear Model (GLM).

9 Point out any two criteria for deciding the number and positions of knot in Spline fitting

10 What is the use of adjusted R-square?

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

11 Explain General linear Hypothesis and develop the F-test 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 $H_0: \beta_2 = \beta_3$

12 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	Singular values	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	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	0.030496		0.9964	0.0588		0.0205	0.6645

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

13 Explain the method of detecting multicollinearity using VIF and conditional Index.

14 Explain Graphical method, spearman's Rank correlation method and White's general heteroscdasticity test for detecting heteroscedassticity.

15 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?

16 In a regression model building study, electricity consumption is investigating the effect of the size of a single – family type and the type of air conditioning used in the house. There are four types of air conditioning systems (no air conditioning / window units / heat bump / central air conditioning) are operated. Write down all possible regression equations for the four classes.

17 Explain Box Jenkins methodology of constructing an ARIMA model.

18 Explain the methods of model validation.

ANSWER **ANY TWO** QUESTIONS SECTION C (2 X 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	165	167	180	155	212	175	190	210	200
BP	130	133	150	128	151	146	150	140	148

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) 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

20 b) Based on a sample of size 16, a model is to be built with 4 regressors. Carry out the forward regression method to decide on the significant regressors at each iteration, given the following information. 15

$$SS_{total} = 1810.509, SS_{Res}(X_1) = 843.8, SS_{Res}(X_2) = 604.224, SS_{Res}(X_3) = 1292.923,$$

$$SS_{Res}(X_4) = 589.24, SS_{Res}(X_1, X_2) = 38.6, SS_{Res}(X_1, X_3) = 818.04, SS_{Res}(X_1, X_4) = 49.8$$

$$SS_{Res}(X_2, X_3) = 276.96, SS_{Res}(X_2, X_4) = 579.23, SS_{Res}(X_3, X_4) = 117.1,$$

$$SS_{Res}(X_1, X_2, X_3) = 32.1, SS_{Res}(X_1, X_2, X_4) = 31.98, SS_{Res}(X_1, X_3, X_4) = 33.89,$$

$$SS_{Res}(X_2, X_3, X_4) = 49.2, SS_{Res}(X_1, X_2, X_3, X_4) = 31.91$$

21 a) Define AR(p), MA(q), ARMA(p,q) and order of integration(d) 12

21 b) Explain Spline regression with an example 8

22 a) Bring out any four specific aspects considered in fitting polynomial regression models. 8

22 b) 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 OLS for a model with three regressors: 12

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 $d_L = 0.34$, $d_U = 1.733$