LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

FIRST SEMESTER – **NOVEMBER 2014**

ST 1821 - APPLIED REGRESSION ANALYSIS

Date : 03/11/2014 Dept. No. Max. : 100 Marks
Time : 01:00-04:00

SECTION – A

Answer ALL the following questions (10 x 2 = 20 marks)

1. Specifiy the test for the slope coefficient of a simpe linear model.
2. Give the decomposition of the overall variation in the data in multiple regression analysis.
3. Give the motivation for standardized regression coefficients.
4. Identify the linearizing transformations required to transform the relation $Y = \beta_0 \exp(\beta_1 X)$ and write down the linearized form.
5. State the variance-stabilizing transformation for a Poisson count variable.
6. Write a note on the 'Model Respecification' method of handling multicollinearity.
7. What is a hierarchical model?
8. If it is known that for $X \leq 30$ there is a linear effect while for $X > 30$ there is a quadratic effect of the regressor X on a dependent variable Y, write down the terms in the model equation allowing an intercept term.
9. Point out any two criteria for deciding the number and positions of knots in spline fitting.
10. Briefly explain any one source for autocorrelation in time-series data.

SECTION – B

Answer any FIVE questions (5 x 8 = 40 marks)

11. 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: $\beta_2 = \beta_3$.
12. A model with five records was built where the Y values were 1.7, 3.5, 2.9, 3.1, 2.5 and the data matrix was $\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & -1 & 3 & 2 & -2 \end{bmatrix}$. Compute the vector of residuals by computing the 'Hat' matrix.
13. Explain the Box-Cox class of power transformations and state the appropriate form(s) relevant for model comparison. Describe the practical method of choosing the power.
14. Explain four sources of multicollinearity.
15. Bring out any four specific aspects considered in fitting polynomial regression models.
16. Explain the tests of significance of autocorrelation coefficients in a time-series.
17. Describe 'Unit Root Test' for stationarity of a time-series.
18. In building a model with four regressors, the singular-value analysis and variance-decomposition proportions were carried out to detect multicollinearity. The following is part of the output obtained in the analysis. Fill up the missing entries and identify the variables that are entangled in collinear relationship:

Eigen	Singular	Condition	Variance Decomposition Proportions

Value (of **X'X**)	value (of **X**)	Indices	Intercept	X_1	X_2	X_3	X_4
	3.575644	1.000000	0.1255	0.1290		0.0207	0.0091
3.58912		1.887384	0.0422	0.0475	0.0008		0.1673
1.56324	1.250296			0.0010	0.0083	0.6023	0.1046
0.01738		27.12248	0.1923	0.1632	0.0041	0.0398	
	0.073417	48.70347	0.6328		0.7842	0.3366	0.0541

SECTION – C

Answer any TWO questions (2 x 20 = 40 marks)

19. (a) Depict five different scenarios that can show up in plotting residuals versus the fitted values and explain how these plots help in detecting model inadequacies.
(b) Give the motivation for the PRESS residuals and define the PRESS statistic .
(15+ 5)

20. (a) Present the formulation of a regression model with non-spherical disturbances and obtain the GLS estimate of the regression parameters. Also, present the ANOVA. Discuss Weighted Least squares as a particular case.
(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. Write the explicit equations for the four classes.
(10+ 10)

21. (a) Describe the 'Forward Model Building' algorithm clearly specifying the partial-F statistics and tests applied..
(b) SS_{total} = 543.152, $SS_{Res}(X_1)$ = 181.266, $SS_{Res}(X_2)$ = 387.88, $SS_{Res}(X_3)$ = 176.774, $SS_{Res}(X_4)$= 253.138, $SS_{Res}(X_1, X_2)$ = 83.088, $SS_{Res}(X_1, X_3)$ = 173.776,$SS_{Res}(X_1, X_4)$ =11.582, $SS_{Res}(X_2, X_3)$ = 35.148, $SS_{Res}(X_2, X_4)$ = 245.414, $SS_{Res}(X_3, X_4)$ = 14.952, $SS_{Res}(X_1, X_2, X_3)$ = 14.762, $SS_{Res}(X_1, X_2, X_4)$ = 9.622, $SS_{Res}(X_1, X_3, X_4)$ = 9.594, $SS_{Res}(X_2, X_3, X_4)$ = 10.168, $SS_{Res}(X_1, X_2, X_3, X_4)$ = 9.572
(5+ 15)

22. (a) Explain Non-parametric regression through 'Kernel Smoothing' and list out any two kernel functions.

(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 residulas obtained by OLS for a model with 3 regressors:
4.818, -10.364, 4.454, -0.727, 4.091, -1.092, -6.272, 3.546, 8.364, -6.818
The relevant DW bounds are given to be d_L = 0.34, d_U = 1.733. (8 +12)