LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

FIRST SEMESTER – NOVEMBER 2017

17/16PST1MC02/ST1821 - APPLIED REGRESSIONANALYSIS

Date: 04-11-2017 Dept. No. Max. : 100 Marks

Time: 01:00-04:00

SECTION – A

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

1. Give the expressions for the OLS estimators of the intercept and slope parameters of a simple linear model with a single regressor.
2. A simple linear model (with one regressor and without intercept) is built using 22 records. The Total Sum of Squares is 367.5 and Residual Sum of Squares is 315. Compute the F-statistic for testing the significance of the slope coefficient.
3. In building a model with an intercept term and 2 regressors using 21 observations, R^2 is found to be 0.73. Find the Adjusted R^2 for the model.
4. Give the motivation for standardized regression coefficients.
5. Define PRESS residuals bringing out their primary use.
6. Identify the linearizing transformations for the relation $Y = X / (\beta_0 X - \beta_1)$.
7. State any two ill-effects of multicollinearity.
8. Distinguish between hierarchical and non-hierarchical polynomial models.
9. Explain the AR(1) scheme for the error terms of a regression model.
10. Describe a 'Random Walk Without Drift'.

SECTION – B

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

11. Discuss the t-test and ANOVA approach to test for significance of the slope coefficient in a simple (single regressor) regression model with intercept.
12. Consider the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$ and suppose that one wishes to test H_0: $\beta_1 = \beta_3$, $\beta_2 + 2\beta_5 = 0$, $\beta_2 = 0$. Discuss this problem under the General Linear Hypothesis framework explicitly stating the 'Reduced Model' equation, the degrees of freedom for Sum of Squares due to the Hypothesis and the F Statistic to test H_0 and the distribution of the Statistic.
13. In building a regression model with an intercept using five observations and a single regressor, the transpose of the data matrix is:
    $\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & -1 & -2 & 1 & 0 \end{bmatrix}$.
    If the residuals are 0.13, -0.34, 0.27, -0.08, 0.02, compute the PRESS residuals.
14. Explain Box-Cox class of power transformations and give the appropriate form(s) required for model comparison. Discuss the method of choosing the power.
15. Explain any four sources of multi collinearity problem. Point out its ill effects.
16. Describe Cubic Spline fitting and discuss the issues in the approach.
17. Discuss the 'Runs Test' in the context of time-series residual analysis. [P.T.O]

18. In building a model with four regressors, the singular-value analysis and variance-decomposition proportions were carried out to detect multicollinearity by standardizing the variables. 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 Value (of **X'X)**	Singular value (of **X)**	Condition Indices	Variance Decomposition Proportions X_1	X_2	X_3	X_4
?	1.725	?	0.1835	0.0317	?	0.0038
?	?	1.7157	0.0994	0.1572	0.1328	?
?	0.1095	?	?	0.1837	0.3097	0.2538
0.001	?	?	0.1543	?	0.1294	0.7302

SECTION – C

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

19. (a) Explain 'Partial Regression Plots' and their uses. State any two cautions to be exercised in using these plots.
(b) Discuss the different approaches to detection of outliers using different versions of residuals of a linear regression model. (10+10)

20. (a) Give the motivation for 'Generalized Least Squares' and discuss the estimation of the regression parameters and ANOVA. Discuss WLS and the issues related to using WLS.
(b) An investigator fits a model for a response variable Y allowing the possibility of different intercepts and different slopes of a single IDV X_1 for three different nationalities (A, B, C) and also for gender differences among males & females. Write down the model equation with appropriate coefficients. Also, specify the intercepts and the slopes for the six classes of individuals. (10 +10)

21. (a) Describe the 'Backward Elimination' algorithm of model building clearly specifying the partial-F statistics and tests applied.
(b) Carry out the 'Backward Elimination' 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} =678.9409, $SS_{Res}(X_1)$ = 220.9667, $SS_{Res}(X_2)$ = 484.8501 , $SS_{Res}(X_3)$ = 226.5841, $SS_{Res}(X_4)$ =316.4217, $SS_{Res}(X_1, X_2)$ =43.9345, $SS_{Res}(X_1, X_3)$ = 217.22, $SS_{Res}(X_1, X_4)$ = 18.6905, $SS_{Res}(X_2, X_3)$ = 103.8607, $SS_{Res}(X_2, X_4)$ = 306.768, $SS_{Res}(X_3, X_4)$ = 14.4761, $SS_{Res}(X_1, X_2, X_3)$ = 18.4536, $SS_{Res}(X_1, X_2, X_4)$ = 12.709, $SS_{Res}(X_1, X_3, X_4)$ = 11.9932, $SS_{Res}(X_2,X_3,X_4)$ =12.0276, $SS_{Res}(X_1,X_2,X_3,X_4)$ =11.9659
(5 +15)

22. (a) Discuss any three causes for occurrence of autocorrelation.
(b) In building a model, the OLS residuals are : 0.12, -0.35, 1.24, 2.05, 1.68, -2.03, 0.87, -0.63, -1.41, 0.06, -0.73, 0.03, -1.38, 0.55, -0.07. Carry out the Durbin-Watson test at 5% significance level and draw your conclusion on the presence of autocorrelation [It is given that d_L =1.08 ,d_U = 1.36] (6+14)

$$$$$$$$$