LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

FIRST SEMESTER – **NOVEMBER 2018**

16/17/18PST1MC02 – APPLIED REGRESSIONANALYSIS

Date: 27-10-2018 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 rationale for the 'non-stochasticity' assumption on the regressor variable in a simple linear model.
2. Explain the 'Ordinary Least Squares' criterion for estimation of regression parameters in a simple linear model.
3. If the R^2 of a model with 3 regressors built with 25 observations is 0.65, find the Adjusted R^2 of the model.
4. Define PRESS residuals and express these in terms of the usual raw residuals explaining the notations.
5. Identify the transformations required to transform the relation $Y = \beta_0 X^{\beta_1}$ to linear form.
6. Specify the variance stabilizing transformations required if V(Y) **α** E(Y) [1 – E(Y)] and the context wherein this relationship arises.
7. Briefly explain any two model-evaluation criteria.
8. Write a note on 'Kernel Smoothing' in non-parametric regression.
9. State any two reasons for occurrence of autocorrelation problem in regression models.
10. State the GLS estimate of the slope coefficient β_1 in the model $Y_t = \beta_0 + \beta_1 X_t + U_t$ where the error terms U_t follow the AR(1) scheme.

SECTION – B

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

11. Under usual notations, develop the 'interval predictor' of a new observation in a simple regression model.
12. A linear regression model with an intercept term and 4 independent variables was built using 100 observations. It was reported that $\Sigma Y_i = 540$, $\Sigma Y_i^2 = 8100$, $\Sigma \hat{Y}_i^2 = 6750$. Construct the ANOVA table and carry out the test for the overall significance of the model.
13. Give the motivation for 'Standardized Regression Coefficients' and explain any one scaling method in this context.
14. Present any four scenarios when residuals are plotted against the fitted values and describe what they indicate.
15. Explain 'Generalized Least Squares' and discuss the estimation of the regression parameters and ANOVA.
16. In building a model with five 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 by using standardized form of the regressors. Fill up the missing entries and identify

the variables that are entangled in collinear relationship:

[Cont'd]

Eigen Values (of X'X)	Singular values (of X)	Condition Indices	Variance Decomposition Proportions				
			X1	X2	X3	X4	X5
?	1.6540	?	0.0012	0.0538	0.0328	0.0002	?
?	?	1.4346	0.0007	?	0.0074	0.0037	0.0027
0.8355	?	?	0.1017	0.0023	0.2021	?	0.1168
?	0.3137	?	?	0.1428	0.0194	0.0068	0.4021
?	0.0361	45.8727	0.6618	0.3999	?	0.9082	0.0262

17. Describe 'Cubic Spline' fitting and the issues in fitting such models.

18. The residuals from a model arranged in time-order had the following signs:
+ + + + – – – – – – – – + – – + + – – – + + + – – + + – – – +.
Carry out the Runs Test for randomness of the sequence.

SECTION – C

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

19. (a) Discuss the context for 'Partial F-test' and 'Extra Sum of Squares method' clearly explaining the notations.
(b) Consider the model $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6$ and suppose that one wishes to test H_0: $\beta_1 = \beta_4$, $\beta_3 - 2\beta_5 = 0$, $\beta_6 = 0$. Discuss this testing 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. (10+10)

20. (a) Explain the Box-Cox class of power transformations and give the appropriate form(s) required for model comparison. Discuss the analytical method of choosing the power.
(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 four different academic specializations (Humanities, Science, Engineering, Medicine) and also for gender differences among men and women. Write down the explicit model equation with appropriate coefficients. Also, specify the intercepts and the slopes for the eight classes of students.
(10 +10)

21. (a) List down four sources of multicollinearity with illustrations.
(b) Carry out 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} = 237.629, $SS_{Res}(X_1)$ = 169.698, $SS_{Res}(X_2)$ = 79.303, $SS_{Res}(X_3)$ = 110.749, $SS_{Res}(X_4)$ = 77.340, $SS_{Res}(X_1, X_2)$ = 36.351, $SS_{Res}(X_1, X_3)$ = 107.368, $SS_{Res}(X_1, X_4)$ = 15.377, $SS_{Res}(X_2, X_3)$ = 5.068, $SS_{Res}(X_2, X_4)$ = 76.027, $SS_{Res}(X_3, X_4)$ = 6.542, $SS_{Res}(X_1, X_2, X_3)$ = 4.211, $SS_{Res}(X_1, X_2, X_4)$ = 6.458, $SS_{Res}(X_1, X_3, X_4)$ = 4.449, $SS_{Res}(X_2, X_3, X_4)$ = 4.198, $SS_{Res}(X_1, X_2, X_3, X_4)$ = 4.188 (5+ 15)

22. (a) Develop the Durbin-Watson test for autocorrelation.
(b) Explain the AR(p), MA(q), ARMA(p,q), I(d), ARIMA(p,d,q) processes. (8+12)

* * * * * *