ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27
M.Sc. STATISTICS – II SEMESTER
SEMESTER EXAMINATION: JULY 2022
ST – 8421: Linear models And Regression Analysis

Time: 2½ Hours   **Max: 70 Marks**

*This question paper contains **ONE** printed page and **TWO** parts*

PART A

I   **Answer any SIX from the following**   **3 x 6 = 18**

1. Obtain the least square estimator of β in the linear model. Is the least square estimator of β unbiased?
2. When do you say that a linear parametric function is estimable?
3. Define R^2 and adjusted R^2 statistics. How they are useful?
4. Describe the construction of Q-Q plots.
5. Discuss F-test in regression analysis.
6. Give a point predictor and prediction interval for a future observation in a multiple linear regression model.
7. Explain generalized least square method of estimation.
8. Discuss the need for data splitting.

PART B

II   **Answer any FOUR from the following**   **13 x 4 = 35**

9. A) State and prove Gauss Markov theorem.
   B) Define general linear model, by stating the assumptions. Also derive the test procedure to test H0: K'β = m.   (6+7)
10. A) Describe less than full rank model. Show that SSR and SSE are independently distributed for less than full rank model.
    B) Define multiple linear regression models. Discuss estimation of parameters in multiple linear regression models.   (8+5)
11. A) Explain forward and backward selection procedures in regression analysis.
    B) Outline graphical analysis of residuals to check the assumption of normality.
    C) Describe Box-Cox method of transformation.   (6+3+4)
12. A) Define the hat matrix H. Show that H and (I-H) are symmetric and idempotent matrices.
    B) Define heteroscedasticity and describe a test for detecting heteroscedasticity.
    C) Define Cook's measure. How does this provide a measure of influence?   (5+5+3)
13. A) Define autocorrelation. Describe Durbin Watson test for autocorrelation. State the assumptions made. What is the limitation of this test procedure?
    B) What are dummy variables in regression analysis?
    C) Define Mallow's Cp- statistic. How it is useful?   (7+3+3)
14. A) Explain the problem of multicollinearity. What are its consequences?
    B) Describe i) Ridge regression.
       ii) Logistic regression   (3+10)

****