LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

B.Sc. DEGREE EXAMINATION - **STATISTICS**

FIFTH SEMESTER – NOVEMBER 2015

ST 5509 - REGRESSION ANALYSIS

Date : 05/11/2015 Dept. No. Max. : 100 Marks

Time : 09:00-12:00

PART-A

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

1. What is a residual?
2. Distinguished between good data collection and poor data collection?
3. Write down multiple regression models and describe each term .
4. Write down the formula for R^2 ?
5. What are the assumptions used for the error term in a regression model?
6. Write down the primary source of multicollinearity ?
7. Write down the linear form of $Y = X/(ax-b)$.
8. What is meant by dummy-variable trap?
9 What is meant by variable selection problem?
10. Define mean absolute percent error?

PART-B

Answer any **FIVE** questions: (5x8=40 marks)

11. Derive the least square estimates of β in a simple linear regression model?
12. Explain the residual plots in detail.
13. Explain the procedure to find an outlier and how to delete from the data.
14. Describe how the dummy variable used as interaction effect, intercept and slope shifter.
15. Explain hypothesis testing of overall significance of the model.
16. Describe the standard error and confidence interval for multiple linear regression models.
17. Fit a straight line to the following data using least square method and estimate the sales for the Year 2016.

Year	2008	2009	2010	2011	2012	2013	2014	2015
Sale of bike (in lakes)	6	6.1	5.2	5	4.6	4.8	4.1	6.2

18. An incomplete ANOVA table for a regression model $Y= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ with n=20 is given below.

Source	SS	DF	MSS	F-ratio
Regression	-	-	-	165.21
Residual	166402.65	-	-	
Total	-	-		

a) Complete the above table. b) Find the R^2.

PART-C

Answer any **TWO** questions: (2x20=40 marks)

19. a) What is meant by heteroscedasticity and how remove from the data set.

    b) Explain PP Plots & QQ Plots.

20. a) What are the sources of multicollinearity? Explain the different methods of diagnosing the problem of multicollinearity.

    b) Explain the problems involved on constructing a model.

21. Fit a regression model $Y= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ for the data given below and also find complete ANOVA table?

Observation number	Delivery time in min(y)	Number of cases (X_1)	Distance in feet (X_2)
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210

22. a) Define : Mean percentage error , Mean Absolute Percentage Error and write the uses.

    b) Find all the errors for following data:

Period	Observed y	Forecast $\hat{y}$
Spring 2010	112	112
Autumn 2010	132	112
Spring 2011	129	124
Autumn 2011	135	124
Spring 2012	142	131
Autumn 2012	124	137
Spring 2013	141	129
Autumn 2013	138	136
Spring 2014	140	137
Autumn 2014	135	139
Spring 2015	140	137
Autumn 2015	144	137

$$$$$$$