LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

B.Sc. DEGREE EXAMINATION – **STATISTICS**

FIFTH SEMESTER – **APRIL 2024**

UST 5503 – REGRESSION ANALYSIS

Date: 20-04-2024 Dept. No. Max. : 100 Marks

Time: 09:00 AM - 12:00 NOON

	SECTION A - K1 (CO1)	
	Answer ALL the Questions	**(10 x 1 = 10)**
1.	**Define the following**	
a)	MAE	
b)	Anderson Darling test	
c)	Hat matrix	
d)	Dummy variable	
e)	VIF.	
2.	**MCQ - Choose the correct option**	
a)	**Interpret the regression equation Y = 0.4 - 0.6X** (i) Expected value of Y is increased by 0.6 units for each one unit increase in X (ii) Expected value of Y is decreased by 0.6 units for each one unit increase in X (iii) Expected value of Y is increased by 0.4 units for each one unit increase in X (iv) Expected value of Y is decreased by 0.4 units for each one unit increase in X	
b)	**What information is contained in the coefficient of determination?** (i) The coefficient of correlation is larger than one **(ii)** Whether r has any significance (iii)We should not partition the total variation (iv)The proportion of total variation in Y that is explained by X	
c)	**Which of the following metrics can be used for evaluating regression models? i) R Squared ii) Adjusted R Squared iii) F Statistics iv) MSE** (i) ii and iv (ii) i and ii (iii) ii, iii and iv (iv) i, ii, iii and iv	
d)	**Influential observation is one which has** (i) Unusual x coordinate (ii) Unusual y coordinate (iii) Unusual x and unusual y coordinate (iv) Usual x and usual y coordinates	
e)	**Specificity is defined as** (i) TP/(TP+FN) (ii) TN/(TN+FP) (iii) FN/(TP+FN) (iv) FP/(TN+FP	
	SECTION A - K2 (CO1)	
	Answer ALL the Questions	**(10 x 1 = 10)**
3.	**Match the following**	
a)	property of least square fit	- $1-(SS_{Res} / SS_T)$
b)	Outlier	- log(p/1-p)
c)	R^2	- Correlation of regressor variables
d)	Multicollinearity	- Large studentized residual

e)	Logit Transformation - $\Sigma e_i x_i = 0$
4.	**TRUE or FALSE** **(5 x 1 = 5)**
a)	When the Sum of squares due to residuals = Sum of squares due to total then $R^2 = 1$
b)	All the residuals are contained in a horizontal band centered around e=0, then the homoscedasticity assumption is violated
c)	When there are k-1 regressor in Multiple Linear Regression model with n observations, then the d.f for residual is n-k
d)	Strong multicollinearity between the regressors result in large variance and covariance of regression coefficients.
e)	Logistic regression is based on Maximum Likelihood Estimation
	SECTION B - K3 (CO2)
Answer any TWO of the following	**(2 x 10 = 20)**
5.	Carryout the test procedure to test the slope coefficient in SLRM
6.	Develop Kolmogrov Smirnov test procedure
7.	Identify the effect of multicollinearity
8.	Obtain estimates of the model parameter in logistic regression
	SECTION C – K4 (CO3)
Answer any TWO of the following	**(2 x 10 = 20)**
9.	Discover 100(1-α)% confidence interval for regression coefficients of SLRM when sigma square is known.
10.	How to test error term involved in a regression model follows a normal distribution
11.	Carryout the test procedure for individual Regression coefficient in Multiple Regression Model
12.	Explain in detail how to handle interaction terms involving dummy variables
	SECTION D – K5 (CO4)
Answer any ONE of the following	**(1 x 20 = 20)**
13.	Evaluate estimators are BLUE in Simple Linear Regression Model
14.	(i) Describe the Multiple linear regression data model using matrix approach. (ii) Explain the model validation in logistic regression analysis
	SECTION E – K6 (CO5)
Answer any ONE of the following	**(1 x 20 = 20)**
15.	The data gives reaction time of 10 men of various ages to a visual stimulus (a flashing light) in a psychological experiment, X is age (in years) and Y is the reaction time (in milliseconds) X: 37 35 41 43 42 50 49 54 60 65 Y: 190 197 205 210 218 226 228 230 234 240 (i) Develop a model $Y = \beta_0 + \beta_1 X + \varepsilon$ to the data (ii) Construct the coefficient of determination and comment on it.
16.	Discuss in detail ROC analysis. How to obtain optimum threshold value.

$$$$$$$$$$$$$$$