

**M. SC. (APPLIED STATISTICS)
(MSCAST)**

**Term-End Examination
June, 2024**

MST-017 : APPLIED REGRESSION ANALYSIS

Time : 3 Hours

Maximum Marks : 50

Note : (i) *Question No. 1 is compulsory.*

(ii) *Attempt any **four** questions from the remaining question nos. 2 to 6.*

(iii) *Use of scientific calculator (non-programmable) is allowed.*

(iv) *Symbols have their usual meanings.*

1. State whether the following statements are True or False. Give reasons in support of your answers : 5×2=10

(a) The model $Y = \beta_0 X^{\beta_1} + \epsilon$ is a linear model.

(b) The full form of the term PRESS is the “predicted response estimated sum of squares” in regression modelling.

- (c) We use Cook's distances to identify the autocorrelation of the error terms.
- (d) If the coefficient of determination for the auxiliary regression model of the explanatory variable X_1 is 0.8, the VIF for X_1 will be 5.
- (e) For a logistic regression model :

$$\hat{\pi} = \frac{\exp(1.5 + 2X)}{1 + \exp(1.5 + 2X)},$$

the odds ratio will be 5.44.

2. The data on the annual yield (in kg/hectare) of a crop and annual rainfall (in cm) are given as follows :

Yield	Rainfall	Yield	Rainfall
21.4	134	22.1	131
23.1	121	23.4	119
20.4	142	22.8	128
24.0	120	21.6	134
22.4	131	23.3	124
23.3	125	21.8	139

If it is given that $\hat{\beta}_0 = 39.3$, $\hat{\beta}_1 = -0.13$, $MSS_{Res} = 0.14$, $SS_x = 614$, then obtain prediction internals and plot the prediction interval along with the given data. (Given that : $t_{(10), 0.025} = 2.228$) 10

3. (a) Explain the normal probability plot. 5
 (b) The weekly wages (in ₹ '000) and working experience (in years) of 20 employees are given in the following table : 5

S. No.	Wages	Experience	S. No.	Wages	Experience
1	72	3	11	86	9
2	71	5	12	88	10
3	73	6	13	69	4
4	72	6	14	86	11
5	85	9	15	85	9
6	78	7	16	87	10
7	73	6	17	83	8
8	79	7	18	77	7
9	79	7	19	82	8
10	82	8	20	71	3

Draw a scatter plot to visualize simple linear, quadratic and cubic regression. Compare them and interpret the results.

4. (a) In a study, the following data on a response variable (Y) and three explanatory variables (X_1 , X_2 and X_3) are recorded :

Y	X_1	X_2	X_3
10	4	12	14
11	7	21	19
12	8	33	33
12	16	48	49
14	21	63	65
15	26	78	80
16	32	96	93
17	38	114	110
18	40	120	122
19	42	126	126

- (i) Check the presence of multicollinearity with the help of correlation matrix. 3
- (ii) Also show the presence of multicollinearity using determinant of matrix. 3

- (b) Describe the issue of autocorrelation in regression modelling along with its various sources. 4
5. (a) For a logistic model fitted on 100 observations, the following values are compacted : 6
- $$(\log L)_F = -64.7 \text{ and } (\log L)_N = -66.8$$
- (b) Compute the McFadden, Cox and Snell and Nagel Kerke pseudo R-squared. Describe the profit and complementary log-log models in brief. 4
6. Write short notes on the following : 4+4+2
- (i) Akaike Information Criterion
 - (ii) Bayesian Information Criterion
 - (iii) Predictive R^2