

**INDIAN STATISTICAL INSTITUTE
CHENNAI CENTER
M.STAT First Year
2014-15 Semester II**

Regression Analysis
Final Examination

Date: 24th April, 2015

Duration: 3 Hours

ANSWER QUESTION NO. 1 AND ANY FIVE QUESTIONS FROM THE REST. TOTAL MARKS IS 100.

1. Write short notes on any five of the following:

- a) High break-down point estimators for robust regression.
- b) Methods for detecting multicollinearity.
- c) Methods for dealing with multicollinearity.
- d) Log-likelihood functions for Poisson and Binary Logistic regression model.
- e) Cochran-Orcutt procedure for eliminating autocorrelation.
- f) LOESS regression procedure.
- g) Sensitivity, specificity and ROC curve in binary logistic regression.

[5*5 = 25]

2. Consider the binary logistic regression model $\text{logit}(p_i) = 25.241 - 2.385X_i$, where p_i denotes the probability of success of an event.

- (a) Compute the probability of failure of the event at $x = 9.5$.
- (b) An observation in the dataset is recorded as $y = 0$ for $x = 10.3$. Compute Pearson residual and Deviance residual for this observation.
- (c) Compute Odds ratio and interpret.

[15]

3. Consider the following data on number of mistakes in various documents of different sizes. Model the number of mistakes as a function of number of pages using Poisson regression.

No. Of mistakes	3	0	1	1	2	5	1	4	5	5	7	3
No. Of Pages	70	54	82	57	66	88	60	95	124	90	130	76

- a) Compute the expected number of mistakes for a document of 100 pages.
- b) Compute the rate ratio and interpret.

[15]

4. Consider the following data on time (X) and oxygen demand (Y). Fit an appropriate non-linear model. Mention the starting values of parameters of the chosen non-linear model.

X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	11.9	16.1	20.0	24.0	27.0	29.5	33.0	35.0	36.2	38.8	39.3	40.8	43.0	45.0	46.1

[15]

5. Find the Robust regression (Huber's t as 1.345) model for the following data:

Observation	1	2	3	4	5	6	7	8	9	10
x	2.0	7.0	14.0	15.0	16.0	20	22.0	23.0	24.0	29.0
y	5.0	2.5	2.6	3.9	5.1	9.5	13.0	14.2	5.2	2.9

[15]

6. Find either the ridge regression (ridge coefficient as 0.04) model or the principal component regression model for the following dataset.

Sl No	Y	A	B	C	Sl No	Y	A	B	C
1	40	135	7.5	42	6	31	142	13.5	56
2	36	140	9.0	46	7	35	140	17.0	63
3	42	145	11	51	8	25	145	23.0	66
4	48	160	13.5	59	9	16	95	5.3	32
5	50	172	17.0	68	10	18	110	7.5	38

[15]

7. Consider the following ordinal logistic regression models for a response variable with three ordered categories (1, 2, 3) and two independent continuous regressor variables X_1 and X_2 .

$$\text{logit } 1 = -20.5 + 0.9x_1 - 0.5x_2$$

$$\text{logit } 2 = -15.2 + 0.9x_1 - 0.5x_2$$

- For a given value of $x_1 = 25$ and $x_2 = 10$, compute the probabilities of Y taking the values of 1, 2 and 3.
- If the observed y value is 2 for the x_1 and x_2 values given in a), then compute Pearson and Deviance residuals.
- Compute the odds ratios for each variable and interpret them.

[15]
