

INDIAN STATISTICAL INSTITUTE,
Chennai Centre
M.Stat. : 2014-16
(Year I - Semester II)

Mid-Semestral Examination: Regression Analysis

Date: 25th February, 2015

Duration: 3½ Hours

ANSWER ALL QUESTIONS. MAXIMUM YOU CAN SCORE IS 100.

1. Consider the simple linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$.
- a) Find the variance-covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$. Find the correlation coefficient between $\hat{\beta}_0$ and $\hat{\beta}_1$.
- b) Show that an alternate formula for the regression sum of squares is

$$SS_R = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2$$

- c) Suppose that a model has been fit to $n = 20$ observations and $R^2 = 0.80$. Test for significance of regression at $\alpha = 0.05$.

[5+5+5=15]

2. A study was conducted at a large engineering farm to examine the relationship between the number of active projects (X) and the number of man-hours (Y) required per week, using the firm's data for the preceding year. A simple linear regression model was fit to the data. The results obtained using $n = 30$ showed that

$$\hat{\beta}_0 = 1.2, \hat{\beta}_1 = 3.4, \bar{Y} = 9.6, se(\hat{\beta}_1) = 0.4 \text{ and } \hat{\sigma}_e = 3.5$$

- i) Test for significance of regression.
- ii) Find 95% prediction interval for $X = 3$.
- iii) For $X = 4$, Y was observed as 20 man-hours. Compute studentized residual and Cook's D for this observation.

[5+5+5=15]

3. Consider the multiple regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.

- a) We have the following summary statistics for a set of data:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 10 & 76.7 & 20.5 \\ 76.7 & 593.81 & 158.01 \\ 20.5 & 158.01 & 51.83 \end{bmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 80.28 \\ 628.972 \\ 160.367 \end{bmatrix} \quad \sum_i y_i^2 = 685.2812$$

Find the regression model and test for its significance at $\alpha = 0.05$.

- b) Suppose that we perform unit length scaling on both the response and regressor variables. We then regress the transformed response variable Y^* on the transformed regressor variables W_1 and W_2 . Find the standardized regression coefficients of the transformed model.
- c) Using the standardized regression coefficients obtained above, find the regression coefficients of the original model.
- d) An observation in the above dataset is recorded as $y = 8.75$ for $x_1 = 8.4$ and $x_2 = 2.8$. Find the fitted response value. Also find the studentized residual and Cook's D for this observation.

[10 + 8 + 7 + 10 = 35]

4. Find an appropriate model for the following data:

X	1	2	3	4	5	6	7	8	9	10
Y	5	50	500	5000	20000	40000	30000	10000	2000	150

[15]

5. Develop an appropriate model using a cubic spline for the following data set:

X	1	2	3	4	5	6	7	8	9	10
Y	50	80	90	60	30	5	30	60	150	400

[15]

6. Develop an appropriate model for the following data set:

Sl No.	Y	X1	X2	Sl No.	Y	X1	X2
1	94	10.2	A	6	90.5	10.8	B
2	93	9.8	A	7	98.0	12.1	C
3	95	12.3	A	8	97.5	11.9	C
4	92	11.4	B	9	97.0	10.1	C
5	91	10.6	B	10	97.4	10.9	C

[10]
