INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR
MID SEMESTER EXAMINATION

Date: 18-09-2015 AN | Time: 2 hours | Full Marks: 60
Autumn Semester : 2015-2016 | Department: Mathematics | No. of students: 100
Subject No.: MA31020/MA41025 | Subject Name: Regression and Time Series Models
Course: B.Tech./M.Sc. 3rd Year (AE, AG, AR, BT, CE, CH, EE, HS, IM, MA, MI, MT, NA)

Instructions: Answer all questions. Marks are indicated at the end of each question. Statistical tables may be used. Choose appropriate levels of significance in different problems. Answers should be given with full details and justifications. No queries will be entertained during the examination regarding doubts.

1. Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$. What are its assumptions? Derive least squares estimators of the parameters of the model and show that they are unbiased. Find their variances. Find the error sum of squares and an unbiased estimator of σ^2. Using distributional assumptions derive distributions of estimators of all parameters $(\beta_0, \beta_1, \sigma^2)$ and use these to develop tests of significance and confidence intervals for these. Further develop confidence interval for expected response at a point x^*. Also find prediction interval for a future response at a point $\hat{x}$. **10M**

2. The following table provides the number of hours (x) that eleven students spent on social media during a weekend and their scores (y) on a test taken on the following Monday.

x	y	x	y
5	76	3	73
6	58	8	61
7	65	0	92
7	74	1	85
3	77	2	82
5	68		

Draw the scatter diagram of (x, y). Do you feel a straight line fit will be good? Guess a straight line. Fit a simple linear regression model in (x, y). Find 95% confidence intervals for the parameters $(\beta_0, \beta_1$ and $\sigma^2)$ of the model. Further test for the significance of the regression coefficient in the linear model. Find the coefficient of correlation and also the coefficient of determination. How much regression does the model explain? Does correlation explain the data? Find the expected score when number of hours is 4. Find a 95% confidence interval for the expected score at this point. Find 95% prediction interval for the score when number of hours is 10. **10M**

(PTO)

3. The following table gives final scores (y), I.Q.'s (x_1) and average number of hours (x_2) (spent on studying the subject per week) of 12 students taking a course on "Social Development". Fit a multiple linear regression model. Find 95% confidence intervals for the coefficients of the model and σ^2. Test for the significance for the regression coefficients. Find R^2 and state how much regression is explained by the model. What is your conclusion regarding relative importance of predictor variables? For this, you also fit simple linear regression models between y and x_1; and also between y and x_2. Are they really well fitting models? Summarize your findings and conclusions. **20M**

x_1	x_2	y	x_1	x_2	y
112	9	83	124	10	91
115	6	77	113	9	79
129	14	95	106	5	36
103	4	49	114	7	58
117	8	63	136	8	93
115	12	80	127	3	84

4. The following table gives size (x) of stores (in thousand square ft) and surplus revenue generated (y) (in lakhs of Rs. per week) for 10 randomly selected shopping malls. Show with the help of a scatter diagram that a second degree curve is appropriate for the data and fit this model. Find 95% confidence intervals for the model parameters and test for their significance. Find R^2 and state how much regression is explained by the model. **10M**

x	y	x	y
21	4.08	6.8	1.94
12	3.40	19.6	4.11
25.2	3.51	14.5	3.16
10.4	3.09	25	3.75
30.9	2.92	19.1	3.60

5. The following table gives the percentages $P(x)$ of a chemical that were used up when an experiment was run at temperature $10x$ (x in degrees Celsius).

x	5^0	10^0	20^0	30^0	40^0	50^0	60^0	80^0
y	.061	.113	.192	.259	.339	.401	.461	.551

Show using a scatter diagram that though the plot looks roughly linear, it can be improved if we consider relationship of the form $1-P(x)=c(1-d)^x$. Fit the new model and test for the significance of the parameters. Use this model to estimate $P(35^0)$. **10M**