INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR
MID SEMESTER EXAMINATION

Date: 03-09-2012 FN Time: 2 hours Full Marks: 60
Autumn Semester : 2012-2013 Department: Mathematics
Subject No.: MA31020 Subject Name: Regression and Time Series Models
Course: B.Tech. 3rd Year (AG, BT, HS, IM, MI) No. of students: 46
Instructions: Answer all questions. Statistical tables may be used. Choose appropriate levels of significance in different problems.

1. The following table shows marks in the college entrance test (X) and final scores (Y) after first year of 10 randomly selected second year students of a college.

Student	Entrance marks	Final Score	Student	Entrance marks	Final Score
1	39	65	6	47	89
2	43	78	7	28	73
3	21	52	8	75	98
4	64	82	9	34	56
5	57	92	10	52	75

(a) Draw the scatter diagram. Do you feel a straight line fit will be good? Guess the line.
(b) Fit a simple linear regression model.
(c) Find 95% confidence intervals for β_0 , β_1 and σ^2. Further test for the significance of the regression coefficient.
(d) Find the coefficient of determination. How much regression does the model explain?
(e) Find a 95% confidence interval for the expected score of a student with entrance marks 50. Also determine a 95% prediction interval for the future score of a student with entrance marks 50.

2. In a biological experiment involving the number of trapped beetles (Y) at the distance (X) from their sex-attractant pheromone, a controlled experiment yielded the following data.

Release distance (cm): X	No of beetles trapped: Y
6.25	5, 3, 4, 6
12.5	5, 2, 5, 4
25	4, 5, 3, 0
50	3, 4, 2, 2
100	1, 2, 2, 3

(a) Plot the scatter diagram of the original data. Also plot the scatter diagram of $X^* = \log_e X$ and Y.
(b) Fit a simple linear regression to the appropriate model as decided by scatter diagrams.
(c) Test for the significance of regression. Find the coefficient of determination.
(d) Carry out a lack of fit analysis to test if the fitted model is adequate.

3. In 1957, the Dutch industrial engineer J.R. DeJong proposed the following model for the time it takes to perform a simple manual task as a function of the number of times the task has been practised: $T \approx t\, s^{-n}$, where T is the time, n is the number of times the task has been practised and t and s are parameters depending on the task and individual. Estimate t and s for the following data set:

T	22.4	21.3	19.7	15.6	15.2	13.9	13.7
n	0	1	2	3	4	5	6

Match the fitted model with a scatter plot of the data. Also find the coefficient of determination? How much model explanation does it give?

4. The chlorine residual (pt/million) Y in a swimming pool at various times (in hours) after being cleaned (X) is as given:

X	2	4	6	8	10	12
Y	1.8	1.5	1.45	1.42	1.38	1.36

Fit a curve of the form $Y \approx a\, e^{-b\,X}$. Match the fitted model with a scatter plot of the data. Also find the coefficient of determination? How much model explanation does it give?

5. A new drug was tested on mice to determine its effectiveness in reducing cancerous tumours. Tests were run on ten mice, each having a tumour of size 4 grams, by varying the amount (X) of drug used and then determining the resulting reduction in the weight (Y) of the tumour. The data are as below:

Coded amount of drug (X)	Tumour weight reduction (Y)	Coded amount of drug (X)	Tumour weight reduction (Y)
1	0.50	6	1.60
2	.90	7	1.53
3	1.20	8	1.38
4	1.35	9	1.21
5	1.50	10	0.65

Show with the help of a scatter diagram that a second degree curve is appropriate for the data and fit this model. Test for the significance of the model.

6. In a study on corporate revenues (in Billion Rs.), the following data is collected on profit Y, capital investment (X_1) and advertising expenditure (X_2):

Y	X_1	X_2	Y	X_1	X_2
15	25	4	1	20	0
16	1	5	16	12	4
2	6	3	18	15	5
3	30	1	13	6	4
12	29	2	2	16	2

Fit a multiple regression and test for significance of overall regression and individual regression coefficients. Find R^2. Give your comments on the suitability of the model. Is their multicollinearity in the model?