CM2605: Simulation and Modeling Technique

Tutorial No 02

1. Consider the following simple linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ for a data set (x_i, y_i), $i = 1, 2, \ldots, n$.
   i. Determine the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 respectively.

   ii. Write down the assumptions you make on the error ε.

   iii. Using the following summary information estimate the above model:
   $\sum x = 96$ $\sum y = 26$ $\sum x^2 = 270$
   $\sum y^2 = 18$ $\sum xy = 58$ $n = 40$

   iv. Predict Y when $x = 3.5$.

   v. For what value of x would the mean value of the corresponding Y be zero.

2. (a) The regression model,

$$Y = \beta x + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2)$$

is called regression through the origin since it presupposes that the expected response corresponding to the input level x=0 is equal to 0.

Suppose that (x_i, Y_i), $i = 1, \ldots, n$ is a data set from this model. Determine the least squares estimator $\hat{\beta}$ of β.

(b) In a study on the occurrence of sodium and chloride in surface streams in central Rhode Island, the following data for chloride concentration Y (in milligrams per liter) and roadway area in the watershed x (in percentage) were obtained:

y	4.4	6.6	9.7	10.6	10.8	10.9	11.8	12.1	14.3	14.7	15.0	17.3
x	0.19	0.15	0.57	0.70	0.67	0.63	0.47	0.70	0.60	0.78	0.81	0.78

i. Fit the model $Y = \beta x + \varepsilon$ for the above data set.

ii. Estimate the mean chloride concentration for a water shed that has 1% roadway area.

3. Consider the following simple linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ for a data set $(x_i, y_i), i = 1, 2, \ldots, n$.

   i. Using the following summary information estimate the above model.

   $$\sum x_i = 169 \qquad \sum y_i = 195 \qquad \sum x_i^2 = 3573$$
   $$\sum y_i^2 = 4389 \qquad \sum x_i y_i = 3911 \qquad n = 10$$

   ii. Predict Y when $x = 12$.

   iii. Give an estimate for the mean of Y when X= 10.

4. For the following data the scatter plot shows a linear relationship.

x	0.10	0.16	0.31	0.37	0.37	0.46	0.50	0.50	0.60	0.70
y	0.96	1.10	0.80	0.84	0.77	0.87	0.60	0.87	0.62	0.61
x	0.75	0.80	0.90	1.00	1.07	1.08	1.11	1.30	1.37	1.54
y	0.70	0.41	0.40	0.41	0.45	0.59	0.25	0.25	0.08	0.10

   Write down the liner model with the intercept term. (Use RStudio)

5. Consider the following data set found in Table. It is representative of a production process with Y being the number of hours it takes to produce the specified item number, X.

(*Y*) Hours	**(*X*) Unit Number**
60	5
45	12
32	35
25	75
21	125

   The proposed model is the two-parameter intrinsic linear model:

   $$Y_i = \beta_0 X_i^{\beta_1} + \varepsilon_i$$

   Where the ε_i's are independent normal with constant variance.

   Since the proposed model is non-linear, Answer the following questions by applying the natural log function on both sides.

   1. Using a suitable statistical technique, estimate β_0 and β_1.
   2. Suggest a method to improve the accuracy of the result.

***