Linear Models. List 2

For this assignment you will use the data from the file ch01pr20.txt. Second column contains the number of copiers and the first column contains the time (in hours) needed to maintain these copiers.

1. Plot the data. Is the relationship approximately linear ?

2. Run the linear regression with y = service time and x= number of machines serviced.

   (a) Give the estimated regression equation.

   (b) Give a 95% confidence interval for the slope.

   (c) Describe the results of the significance test for the slope. Give the hypothesis being tested, the test statistic with degrees of freedom, the P-value, and your conclusion in a brief sentence.

3. Give an estimate of the mean service time that you would expect if 11 machines were serviced; and a 95% confidence interval for this estimate.

4. Give a prediction for the actual service time that you would expect if 11 machines were serviced; and 95% prediction interval for this time.

5. Plot the data with the 95% prediction bounds for individual observations.

6. Assume $n = 40$, $\sigma^2 = 120$, $SSX = \sum(X_i - \bar{X})^2 = 1000$.

   (a) Find the power for rejecting the null hypothesis that the regression slope is zero using a $\alpha = 0.05$ significance test when the true slope is $\beta_1 = 1$.

   (b) Plot the power as a function of β_1 for values of β_1 between -2 and 2.

7. Generate the vector $X = (X_1, \ldots, X_{200})^T$ from the multivariate normal distribution $N\left(0, \frac{1}{200}I\right)$. Then generate 1000 vectors Y from the model $Y = 5 + \beta_1 X + \epsilon$, where

   (a) $\beta_1 = 0$, $\epsilon \sim N(0, I)$

   (b) $\beta_1 = 0$, $\epsilon_1, \ldots, \epsilon_{200}$ are iid from the exponential distribution with $\lambda = 1$.

   (c) $\beta_1 = 1.5$, $\epsilon \sim N(0, I)$

   (d) $\beta_1 = 1.5$, $\epsilon_1, \ldots, \epsilon_{200}$ are iid from the exponential distribution with $\lambda = 1$.

   For each replication of the experiment test the hypothesis that $\beta_1 = 0$ and estimate the probability of rejection by the frequency of rejections in your sample (separately for each of the points (a)-(d)). Compare these estimated probabilities to the theoretical probability of the type I error (a and b) and the theoretical power (c and d) calculated under the assumption that the noise ϵ has the normal distribution. Summarize the results.

Problems to be calculated by hand.

8. You use $n = 20$ observations to fit the linear model $Y = \beta_0 + \beta_1 X + \epsilon$. Your estimators are $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 3$ and $s = 4.0$.
   (a) The estimated standard deviation of $\hat{\beta}_1$, $s(\hat{\beta}_1)$, is equal to 1. Construct the 95 % confidence interval for β_1.
   (b) Do you have statistical evidence to believe that Y depends on X ?
   (c) The 95% confidence interval for $E(Y)$ when $X = 5$ is [13,19]. Find the corresponding prediction interval.

Małgorzata Bogdan