LINEAR REGRESSION MODELS W4315

HOMEWORK 2 ANSWERS

October 26, 2011

Instructor: Frank Wood

1. (35 points) Consider a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$. Let $X_i = i \ \forall \ i \in \{1, 2, \ldots, 20\}$ (integers from 1 to 20). We denote the maximum likelihood estimates of β_0 and β_1 as b_0 and b_1 under the normal regression model. Let the true values be $\beta_0 = 2$, $\beta_1 = 0.3$ and $\sigma^2 = 4$.

a. Re-derive the exact sampling distributions of b_0 and b_1. Show your work and state your assumptions.

b. Using matlab, write a script to generate $Y_i = \beta_0 + \beta_1 X + \epsilon \ \forall i$ according to the given model (there are 20 X_i's, so you need to generate 20 Y_i's), and calculate the maximum likelihood estimates b_0 and b_1 based on the $\{X_i, Y_i\}$ data. Repeat this process 1,000 times to generate 1,000 estimates of β_0 and β_1. For each parameter, draw a density histogram of these estimates.

c. Using matlab, superimpose the b_0 and b_1's theoretical probability density functions onto the two histogram plots respectively. Compare the plots.

Answer:

This is the answer to question 1.

2. (25 points) Write a matlab function to produce the ANOVA table as in TABLE 2.2 page 67 of the textbook. Specifically, the interface of your function must be

function [SSR, SSE, SSTO, df_R, df_E, df_TO] = anova_1d(X, Y)

This function accepts X and Y as arguments and returns SSR, SSE, $SSTO$ and their associated degrees of freedom. (As in homework 1, you are only allowed to use basic matlab commands to write this function (not, for instance, *regress*.)

Answer:

3. (40 points) Use the data in the file "problem3.txt" on the course website. This is a 20 by 2 matrix, with the first column being X and second column being Y. Assume a simple linear regression model.

a. Give the ordinary least square estimates of β_0, β_1 and σ^2. Draw a scatterplot of the raw data and overlay the fitted line on it.

b. F-test is used to test the linear relationship between X and Y. Calculate the F-test statistic and p-value. Draw the probability density function of F distribution with appropriate degrees of freedom. What does the p-value mean on the graph?

c. Write your own function to calculate p-value of F-test. The interface of your function must be:

function p_value = p_value_of_F_test(X, Y)

It takes X and Y as arguments and return p-value. Please implement the function defined in "p_value_of_F_test.m" that you can find on the homework section of the course website. We have provided you the exact function interface in that file. (As before, you are only allowed to use basic matlab commands.)

Answer: