

LINEAR REGRESSION MODELS W4315

HOMEWORK 2 QUESTIONS

November 5, 2011

Instructor: Frank Wood

1. (35 points) Consider a simple linear regression model $Y = \beta_0 + \beta_1 X + \epsilon$. X takes value at all the integers from 1 to 20. We denote the ordinary least squared estimates of β_0 and β_1 as b_0 and b_1 . Now assume that $\beta_0 = 2$, $\beta_1 = 0.3$ and $\sigma^2 = 4$.

a. What are the exact distributions of b_0 and b_1 ?

b. Generate Y according to the given model (there are 20 X 's, so you need to generate 20 Y 's), and calculate the OLS estimates b_0 and b_1 based on X and the simulated Y . Repeat the process 1,000 times and you will have 1,000 estimates of β_0 and β_1 . For each parameter, draw a density histogram of the estimates.

c. Superimpose the b_0 and b_1 's probability density functions onto the two histograms respectively. Is the histogram a close approximation of the curve?

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. Please complete the function defined in the file “anova_1d.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 in homework 1, you are only allowed to use basic matlab commands to write this function.)

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.)