LINEAR REGRESSION MODELS W4315

HOMEWORK 1 QUESTIONS

September 6, 2010

Professor: Frank Wood

1. (20 points) Let $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ be a linear regression model with distribution of error terms unspecified (but with mean $E(\epsilon) = 0$ and variance $V(\epsilon_i) = \sigma^2$ (σ^2 finite) known). Show that $s^2 = MSE = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}$ is an unbiased estimator for σ^2. $\hat{Y}_i = b_0 + b_1 X_i$ where $b_0 = \bar{Y} - b_1 \bar{X}$ and $b_1 = \frac{\sum_i((X_i - \bar{X})(Y_i - \bar{Y}))}{\sum_i (X_i - \bar{X})}$

2. (20 points) Derive the maximimum likelihood estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\sigma}^2$ for parameters β_0, β_1, and σ^2 for the normal linear regression model (i.e. $\epsilon_i \sim_{iid} N(0, \sigma^2)$).

3. (10 points) File 'unif.txt' contains 200 numbers randomly generated from a uniform distribution U(2,5). Read these numbers into MATLAB (using 'textread' for instance) and do the following:

a. Take these 200 number as a population. Use command 'randsample' of MATLAB, draw 100 numbers out of this population randomly with replacement and plot one histogram of these 100 numbers.

b. Use command 'rand' in MATLAB, draw 100 samples directly from U(2,5) and plot another histogram of these numbers.

c. Compare the two histograms, what can you say about the difference between the distribution of samples from the population and from the uniform distribution itself? For the convenience of comparison, you may want to overlay two histograms onto one graph and to see if any apparent difference.

4. (10 points) File 'normal.txt' contains 200 randomly generated numbers from Normal distribution N(-1,2^2). Like in problem 3, do the following:

a. Take these 200 number as a population. Use command 'randsample' of MATLAB, draw 100 numbers out of this population randomly with replacement and plot a histogram of these 100 numbers.

b. Use command 'randn' in MATLAB, draw 100 samples directly from N(-1,2^2) and plot a histogram of these numbers.

c. Compare the two histograms, can you get the similar conclusion as that of problem 3?

5. (40 points) Copier maintenance.[1] The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair services on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive number of minutes spent by the service person. Assume that first-order regression model($Y_i = b_0 + b_1 X_i + \epsilon_i$) is appropriate.

i:	**1**	**2**	**3**	...	**43**	**44**	**45**
X_i	2	4	3	...	2	4	5
Y_i	20	60	46	...	27	61	77

a. Obtain estimated regression function.
b. Plot the estimated regression function and the data. How well does the estimated regression function fit the data?
c. interpret b_0 in your estimated regression function. Does b_0 provide any relevant information here? Explain.
d. Obtain a point estimate of the mean service time when $X = 5$ copiers are serviced.

Notice: You can get data for this problem on www.mhhe.com/KutnerALRM4e. Use MATLAB, do not use any other programming language. Only basic MATLAB operators are allowed, do not use any built-in functions to do the regression, i.e. the function "regress" cannot be used except, perhaps, to verify that your answer is correct before submitting your own implementation.

[1] This is problem 1.20 in "Applied Linear Regression Models(4th edition)" by Kutner etc.)