Homework 2, Stat 3

1. Consider the following model:

$$y_i = \beta x_i + \epsilon_i x_i, \qquad i = 1, 2, \cdots, n$$

where y_i are observed, x_i are known positive constants and β is an unknown parameter. The errors ϵ_i are independently and identically distributed random variables having the probability density function

$$f(u) = \frac{1}{2\lambda} \exp\left(-\frac{|\, u \,|}{\lambda}\right), \qquad -\infty < u < \infty$$

and $\lambda > 0$ is an unknown parameter.

   (a) Find the least squares estimator of β

   (b) Find the maximum likelihood estimator of β

2. Consider a multiple linear regression model with p predictors and a nonzero intercept with iid normal errors.

   (a) Write down the model explaining all the terms.

   (b) Show that the least squares estimator of β has a multivariate nomal distribution. What are the parameters of this distribution?

   (c) Derive the distribution of the Sum of Squares Residual (SSR).

   (d) Show that the least squares estimator of β and SSR are independent.

   [You may use any result from probability, linear algebra or calculus by preperly stating it]

3. Consider the data in the file seatpos.xlsx

   Data description: Car drivers like to adjust the seat position for their own comfort. Car designers would find it helpful to know where different drivers will position the seat depending on their size and age. Researchers at the HuMoSim laboratory at the University of Michigan collected data on 38 drivers.

   (a) Fit a multiple linear regression of hipcenter on all other variables.

   (b) Comment on the significance of the coefficients and the R^2.

   (c) What do you suspect about collinearity from your comment? Compute the correlation matrix of the predictors to check if your suspicion is correct. Explain your conclusion.

   (d) Fit simple linear regression models of hipcenter on each of the predictors separately. Which predictor gives the highest R^2? Is it very different from the R^2 you obtained from part (b)? Is the regression coefficient significant?

(e) Summarize your findings.

4. Consider the data in the file earthquake.xlsx

   Data description: The earthquake data frame contains measurements of latitude, longitude, focal depth and magnitude for all earthquakes having magnitude greater than 5.8 between 1964 and 1985.

   Produce the diagnostic plots of residuals for the multiple linear regression of magnitude using all other variables as predictors. Comment on the plots.