Tutorial Problems

Question 1

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where $\epsilon_i \sim NID(0, \sigma^2)$.

(a) Prove that the least squares estimates for the parameters satisfy

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}, \quad \text{and}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y}) x_i}{\sum_{i=1}^n (x_i - \bar{x}) x_i}.$$

(Lecture 3&4)

(b) Show that the residuals, $R_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ sum to 0. (Assumed knowledge)

(c) Write the least squares estimator for the intercept in the form $\sum_{i=1}^n a_i Y_i$. (Assumed knowledge)

(d) Show that the least squares estimator for β_0 is unbiased by using c). (Lecture 3&4)

(e) (Lecture 5) Recall that you can write the above model in a matrix notation with $\mathbf{X} = \begin{pmatrix} \mathbf{1}_n & \boldsymbol{x} \end{pmatrix}$. Show that

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

(f) (Lecture 5) Show that

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{nS_{xx}} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}.$$

(g) (Lecture 3-5) Show that using matrix notation or otherwise that

i. $Var(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$,

ii. $Var(\hat{\beta}_0) = \sigma^2 \dfrac{\sum_{i=1}^n x_i^2}{nS_{xx}}$, and

iii. $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \dfrac{\bar{x}}{S_{xx}}$.

Question 2

The leverage for observation i is a useful number between 0 and 1, which can be calculated by $h_{ii} = \boldsymbol{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \boldsymbol{x}_i$. For the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

show that the formula for the leverage is

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}.$$

Note the points with highest leverage are the ones with x coordinate farthest from $\bar{x}$. (Lecture 8)

Computer Problems

Question 1

In an experiment to investigate the amount of a drug retained in the liver of a rat, 19 rats were randomly selected, weighed, placed under light ether anesthesia and given an oral dose of the drug. The dose an animal received was determined as approximately 40mg of the drug per kilogram of body weight, since liver weight is known to be strongly related to body weight and it was felt that large livers would absorb more of a given dose than smaller livers. After a fixed length of time each rat was sacrificed, the liver weighed, and the percent of the dose in the liver determined. The data can be found in `ratliver.txt`

The experimental hypothesis was that, for the method of determining the dose, there is no relationship between the percentage of the dose in the liver (Y) and the body weight (x_1), liver weight (x_2), and relative dose (x_3). Consider the following model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad \epsilon_i \sim NID(0, \sigma^2). \tag{1}$$

(a) Give the fitted least squares multiple regression equation, the residual sum of squares as well as an estimate for the error variance (Lecture 6).

(b) Construct a 90% confidence interval for the error standard deviation (Lecture 6).

(c) Calculate a 95% confidence interval for the expected Y value when $x_1 = 165$, $x_2 = 8.0$ and $x_3 = 0.85$ (Lecture 7).

(d) Can any of the explanatory variables be dropped from the model based on t-tests? (Lecture 6&9)

(e) Determine if there are any high leverage points. (Lecture 8)

(f) Show that there is one outlier (Lecture 8).

(g) Refit the model ignoring the outlier. Determine the square of the multiple correlation coefficient. Comment. (Lecture 6&8)

(h) Calculate 95% confidence intervals for the parameter estimates in the model. Do any of the intervals contain 0? (Lecture 6)

(i) Calculate the sample standard deviation of the y sample ignoring the outlier (assumed knowledge).

(j) By comparing the sample standard deviation of the y to the error standard deviation in Model (1) (ignoring the outlier) does there appear to be any useful information for predicting Y contained in the x variables (new material)?

(k) Finally, perform an F-test for $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ vs. $H_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0 \vee \beta_3 \neq 0$. (Lecture 9)