Multiple Linear Regression

Due: Monday 09/13 (11.00PM)
Submission: On Gradescope

You do not need to use *R* Markdown for this homework. You can write or type your answers and scan/upload a PDF file on Gradescope.

Part I: Practice Questions

You do not need to submit these questions.

1. Setup the $\mathbf{X}$ matrix and β vector for each of the following regression models. Assume $i = 1, \ldots, 4$.
   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$
   (b) $\sqrt{Y_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \varepsilon_i$

2. Compute the trace of the hat matrix.

3. Consider the following regression model

$$\mathbf{Y} = \mathbf{1}_n \beta + \epsilon$$

   Compute the least squares estimator of β and compute the corresponding hat matrix $\mathbf{H}$.

4. For a general linear regression model in which $\mathbf{X}$ may or may not have full rank, show that

$$\sum_{i=1}^{n} \hat{Y}_i (Y_i - \hat{Y}_i) = 0$$

5. Consider the multiple regression model

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \; i = 1, \ldots, n$$

   where ϵ_i are uncorrelated, with $\mathbb{E}(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. Assuming that ε_i are independent normal random variables, state the likelihood function and obtain the maximum likelihood estimators of β_1, and β_2.

Part II: Homework Questions – to be submitted

1. Setup the $\mathbf{X}$ matrix and β vector for each of the following regression models. Assume $i = 1, \ldots, 4$.
   (a) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$

(b) $\sqrt{Y_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \varepsilon_i$

2. Consider the Simple Linear Model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where $i = 1, \ldots, n$, and $\epsilon_i \sim IID(0, \sigma^2)$. Suppose that the value of the predictor x_i is replaced by $cx_i + d$, where c, d are some non-zero constant. **Show** how are $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}^2$, R^2 and the t-test of $H_0 : \beta_1 = 0$ affected by this change. *Justify your answer.*

3. Obtain the maximum likelihood estimators in a simple linear regression model with normal error terms.

4. Show that $Cov(\mathbf{r}) = \sigma^2(\mathbf{I}_n - \mathbf{H})$.

5. Show that if $\mathbf{X}$ has full rank,

$$(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)^T\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta)$$

holds, and hence deduce that the left side is minimized uniquely when $\beta = \hat{\beta}$.