Biomath 204 Homework 1

Benjamin Chu

January 22, 2017

Problem 1. Prove the Gauss-Markov theorem for β_0 in the following simple linear regression model:

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

assuming $E(\epsilon_i) = 0, Var(\epsilon_i') = \sigma^2, Cov(\epsilon_i, \epsilon_j) = 0$.

Proof. In class we derived that $b_0 = \bar{Y} - b_1 \bar{X}$. To show b_0 is unbiased, note:

$$E(\bar{Y}) = \frac{1}{n} E(\sum_i Y_i) = \frac{1}{n} \sum_i [\beta_0 + \beta_1 X_i] = \frac{1}{n} n\beta_0 + \frac{1}{n}\beta_1 \sum_i X_i = \beta_0 + \beta_1 \bar{X}$$

On the other hand, $E(\beta_1 X_i) = \beta_1 \bar{X}$, so

$$E(b_0) = E(\bar{Y} - \beta_1 X_i) = E(\bar{Y}) - E(\beta_1 X_i) = \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} = \beta_0.$$

To show linearity in Y, recall in lecture we showed

$$b_1 = \sum_i k_i Y_i, \quad k_i = \frac{(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}.$$

Using this, we have

$$\begin{aligned}
b_0 &= \bar{Y} - b_1 \bar{X} \\
&= \bar{Y} - \big(\sum_i k_i Y_i\big) \bar{X} \\
&= \frac{1}{n} \sum Y_i - \frac{1}{n} \sum k_i Y_i X_i \qquad (1) \\
&= \frac{1}{n} \sum [Y_i - k_i Y_i X_i] \\
&= \frac{1}{n} \sum Y_i (1 - k_i X_i)
\end{aligned}$$

Finally, to show min variance, note

□