Homework 2

Aiden Kenny
STAT GR5205: Linear Regression Models
Columbia University
October 5, 2020

Question 1 *Collaborators:* None

Supposed for $\mathbf{x}, \mathbf{y}, \boldsymbol{\epsilon}, \mathbf{1} \in \mathbb{R}^n$, where $\mathbf{x}, \mathbf{1}$ are *fixed* vectors and $\mathbf{y}, \boldsymbol{\epsilon}$ are *random* vectors, the simple linear regression model

$$\mathbf{y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{x} + \boldsymbol{\epsilon}$$

holds, with $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\mathrm{Var}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}$. The least-squares estimators are given by

$$\hat{\beta}_1 = \frac{(\mathbf{x} - \bar{x}\mathbf{1})^T(\mathbf{y} - \bar{y}\mathbf{1})}{\|\mathbf{x} - \bar{x}\mathbf{1}\|^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n-2}\left\|\mathbf{y} - \hat{\beta}_0\mathbf{1} - \hat{\beta}_1\mathbf{x}\right\|^2.$$

(a) We first determine several properties of $\mathbf{y}$ (a random vector) and $\bar{y}$ (a random variable). For $\mathbf{y}$, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{y}] &= \mathbb{E}[\beta_0\mathbf{1} + \beta_1\mathbf{x} + \boldsymbol{\epsilon}] = \beta_0\mathbf{1} + \beta_1\mathbf{x} + \mathbb{E}[\boldsymbol{\epsilon}] = \beta_0\mathbf{1} + \beta_1\mathbf{x}, \\
\mathrm{Var}[\mathbf{y}] &= \mathrm{Var}[\beta_0\mathbf{1} + \beta_1\mathbf{x} + \boldsymbol{\epsilon}] = \mathbf{0} + \mathbf{0} + \mathrm{Var}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}.
\end{aligned}$$

That is, for each y_i, we have $\mathbb{E}[y_i] = \beta_0 + \beta_1 x_i$ and $\mathrm{Var}[y_i] = \sigma^2$. We also have $\mathrm{Cov}[y_i, y_j] = 0$ for all $i \neq j$. For $\bar{y}$, we have

$$\begin{aligned}
\mathbb{E}[\bar{y}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n y_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[y_i] = \frac{1}{n}\sum_{i=1}^n \big(\beta_0 + \beta_1 x_i\big) = \beta_0 + \beta_1\bar{x}, \\
\mathrm{Var}[\bar{y}] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n y_i\right] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[y_i] = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Expanding out $\hat{\beta}_1$ gives us

$$\hat{\beta}_1 = \frac{(\mathbf{x} - \bar{x}\mathbf{1})^T(\mathbf{y} - \bar{y}\mathbf{1})}{\|\mathbf{x} - \bar{x}\mathbf{1}\|^2} = \frac{1}{\|\mathbf{x} - \bar{x}\mathbf{1}\|^2}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

For notational ease, we are going to multiply both sides of this estimate by $\|\mathbf{x} - \bar{x}\mathbf{1}\|^2$, since it is just a constant. Taking the expected value of $\|\mathbf{x} - \bar{x}\mathbf{1}\|^2 \cdot \hat{\beta}_1$ yields

$$\begin{aligned}
\mathbb{E}\Big[\|\mathbf{x} - \bar{x}\mathbf{1}\|^2 \cdot \hat{\beta}_1\Big] &= \|\mathbf{x} - \bar{x}\mathbf{1}\|^2 \cdot \mathbb{E}[\hat{\beta}_1] = \mathbb{E}\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})\right] \\
&= \sum_{i=1}^n \mathbb{E}\big[(x_i - \bar{x})(y_i - \bar{y})\big] = \sum_{i=1}^n (x_i - \bar{x})\big(\mathbb{E}[y_i] - \mathbb{E}[\bar{y}]\big) \\
&= \sum_{i=1}^n (x_i - \bar{x})\big(\beta_0 + \beta_1 x_i - \beta_0 - \beta_1\bar{x}\big) = \beta_1\sum_{i=1}^n (x_i - \bar{x})^2 = \beta_1 \cdot \|\mathbf{x} - \bar{x}\mathbf{1}\|^2.
\end{aligned}$$

Dividing both sides of the equation shows that $\mathbb{E}[\hat{\beta}_1] = \beta_1$. Next, taking the expected value of $\hat{\beta}_0$ gives us

$$\mathbb{E}[\hat{\beta}_0] = \mathbb{E}\big[\bar{y} - \hat{\beta}_1\bar{x}\big] = \mathbb{E}[\bar{y}] - \bar{x}\mathbb{E}[\hat{\beta}_1] = \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} = \beta_0.$$

Finally, taking the expected value of $\hat{\sigma}^2$ leads to

(b) Looking at the expanded equation for $\hat{\beta}_1$, we have

$$\begin{aligned}\|\mathbf{x}-\bar{x}\mathbf{1}\|^2\cdot\hat{\beta}_1 &= \sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^n (x_i-\bar{x})y_i - \bar{y}\sum_{i=1}^n (x_i-\bar{x}) \\ &= \sum_{i=1}^n (x_i-\bar{x})y_i - \bar{y}\big(n\bar{x}-n\bar{x}\big) = \sum_{i=1}^n (x_i-\bar{x})y_i.\end{aligned}$$

That is, we are able to remove the $\bar{y}$ from the summation entirely. By taking the variance of $\|\mathbf{x}-\bar{x}\mathbf{1}\|^2\cdot\hat{\beta}_1$, we have

$$\begin{aligned}\operatorname{Var}\Big[\|\mathbf{x}-\bar{x}\mathbf{1}\|^2\cdot\hat{\beta}_1\Big] &= \|\mathbf{x}-\bar{x}\mathbf{1}\|^4\cdot\operatorname{Var}[\hat{\beta}_1] = \operatorname{Var}\left[\sum_{i=1}^n (x_i-\bar{x})y_i\right] \\ &= \sum_{i=1}^n \operatorname{Var}\big[(x_i-\bar{x})y_i\big] = \sum_{i=1}^n (x_i-\bar{x})^2\operatorname{Var}[y_i] = \sum_{i=1}^n (x_i-\bar{x})^2\sigma^2 = \sigma^2\cdot\|\mathbf{x}-\bar{x}\mathbf{1}\|^2,\end{aligned}$$

and dividing both sides by $\|\mathbf{x}-\bar{x}\mathbf{1}\|^2$ shows that $\operatorname{Var}[\hat{\beta}_1] = \sigma^2/\|\mathbf{x}-\bar{x}\mathbf{1}\|^2$. Similarly, taking the variance of $\hat{\beta}_0$ gives us

$$\operatorname{Var}[\hat{\beta}_0] = \operatorname{Var}\big[\bar{y}-\hat{\beta}_1\bar{x}\big] = \operatorname{Var}[\bar{y}] + \bar{x}^2\operatorname{Var}[\hat{\beta}_1] = \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{\|\mathbf{x}-\bar{x}\mathbf{1}\|^2} = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}}{\|\mathbf{x}-\bar{x}\mathbf{1}\|^2}\right).$$

Question 2 *Collaborators:* None

stop