

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 $\text{Var}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}$. 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Question 2 *Collaborators:* None

Letting $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\mathbf{X} = [\mathbf{1} \ \mathbf{x}] \in \mathbb{R}^{n \times 2}$, the simple linear regression model is given by $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The MSE is then given by $Q = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$.

(a) We first expand the MSE to get

$$Q = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}.$$

By differentiating Q with respect to $\boldsymbol{\beta}$ and setting it equal to $\mathbf{0}$, we have

$$\begin{aligned}\frac{\partial Q}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{y}^T \mathbf{y} - 2 \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \stackrel{\text{set}}{=} \mathbf{0},\end{aligned}$$

and solving for $\boldsymbol{\beta}$ gives us $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.

(b) Now suppose that we have p different predictors, each with n observed values that are not all identical. Let $\mathbf{x}_j \in \mathbb{R}^n$ be the vector containing the observations for the j th predictor. Define the matrix $\mathbf{X} = [\mathbf{1} \ \mathbf{x}_1 \ \cdots \ \mathbf{x}_p] \in \mathbb{R}^{n \times (p+1)}$ be the matrix whos j th column is \mathbf{x}_{j-1} (and first column is $\mathbf{1}$). In addition, let $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T \in \mathbb{R}^{p+1}$ be the vector containing $p+1$ scalars. The multiple linear regression model is given by $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$; that is, it has the same form as the simple linear regression model, and we can see that simple linear regression is when $p = 1$. Because of this, the estimated coefficients $\hat{\boldsymbol{\beta}}$ take the same form as before: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.

(c) The fitted values $\hat{\mathbf{y}}$ are given by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

(d) Similar to simple linear regression, the “normal equation” for the multivariate regression setting is found during the derivation of $\hat{\boldsymbol{\beta}}$, and is given by

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}.$$

Letting $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ be the vector of observed residuals, we have $\mathbf{X}^T \mathbf{e} = \mathbf{0}$, meaning $\mathbf{1}^T \mathbf{e} = 0$ and $\mathbf{x}_j^T \mathbf{e} = 0$ for all j . In other words, the residuals sum to zero and the residuals weighted by each predictor sum to 0. Because of this result, we also know that the residuals weighted by the fitted values sum to zero, i.e.

$$\hat{\mathbf{y}}^T \mathbf{e} = (\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y})^T \mathbf{e} = \mathbf{y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e} = \mathbf{y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{0} = \mathbf{0}.$$

The other key results are that $\hat{\mathbf{y}}^T \mathbf{1} = \mathbf{y}^T \mathbf{1}$ and $\bar{y} = \bar{\mathbf{x}}^T \hat{\boldsymbol{\beta}}$, where $\bar{\mathbf{x}} = (1, \bar{x}_1, \dots, \bar{x}_p)^T$ is the vector whos j th entry is the mean of the $(j - 1)$ th variable.

(e) We have

$$\begin{aligned} \|\hat{\mathbf{y}} - \bar{y}\mathbf{1}\|^2 &= (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^T (\hat{\mathbf{y}} - \bar{y}\mathbf{1}) \\ &= \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} + \bar{y}^2 \mathbf{1}^T \mathbf{1} \\ &= \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \bar{y} \mathbf{y}^T \mathbf{1} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} + \bar{y}^2 \mathbf{1}^T \mathbf{1} && \text{(since } \hat{\mathbf{y}}^T \mathbf{1} = \mathbf{y}^T \mathbf{1}) \\ &= \hat{\mathbf{y}}^T (\mathbf{y} - \mathbf{e}) - \bar{y} \mathbf{y}^T \mathbf{1} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} + \bar{y}^2 \mathbf{1}^T \mathbf{1} && \text{(since } \hat{\mathbf{y}} = \mathbf{y} - \mathbf{e}) \\ &= \hat{\mathbf{y}}^T \mathbf{y} - \hat{\mathbf{y}}^T \mathbf{e} - \bar{y} \mathbf{y}^T \mathbf{1} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} + \bar{y}^2 \mathbf{1}^T \mathbf{1} \\ &= \hat{\mathbf{y}}^T \mathbf{y} - \bar{y} \mathbf{y}^T \mathbf{1} - \bar{y} \hat{\mathbf{y}}^T \mathbf{1} + \bar{y}^2 \mathbf{1}^T \mathbf{1} && \text{(since } \hat{\mathbf{y}}^T \mathbf{e} = \mathbf{0}) \\ &= (\hat{\mathbf{y}} - \bar{y}\mathbf{1})^T (\mathbf{y} - \bar{y}\mathbf{1}). \end{aligned}$$

Essentially, we expanded out the equation and used the key results from part (d) to manipulate the equation and get the desired result.

Question 3 *Collaborators:* None

- (a) Our hypothesis test is given by $H_0 : \beta_1 \leq 0$ vs. $H_a : \beta_1 > 0$, and we fail to reject H_0 . The analyst claiming that there is no linear association between X and Y is *incorrect*. The data indicates that $\beta_1 \leq 0$, not $\beta_1 = 1$, meaning there could be a negative linear association.
- (b) While this situation is more subtle, the analyst is again incorrect. In the first situation, we are estimating the true value of $\mathbb{E}[Y|X = x_0]$, so we are trying to estimate an underlying feature of the random variable Y . In the second situation, we are trying to predict the average of *random samples* from Y , i.e. we want to estimate $\mathbb{E}[\bar{Y}|X = x_0]$. One reason for this confusion is that the two results are the same. For example, if $Z \sim N(\mu, \sigma^2)$, then $\bar{Z} \sim N(\mu, \sigma^2/n)$, which means $\mathbb{E}[Z] = \mathbb{E}[\bar{Z}] = \mu$. It is important to note that even though the numerical answer is the same, *what* we are doing in each situation is different.
- (c) At the point $X = x_0$, a 95% confidence interval for Y_0 is given by

$$\hat{Y}_0 \pm t_{0.975, n-2} \sqrt{\text{MSE} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}.$$

However, this confidence interval would *not* be the same at some different point $X = x_h \neq x_0$. At this point, a 95% confidence interval would be given by

$$\hat{Y}_h \pm t_{0.975, n-2} \sqrt{\text{MSE} \left(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}.$$

This new confidence interval is based around \hat{Y}_h , not \hat{Y}_0 , and its width is based on x_h , not x_0 .