STA 302f15 Assignment Five[1]

These problems are preparation for the quiz in tutorial on Friday October 15th, and are not to be handed in.

The general linear regression model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times (k+1)$ matrix of observable constants, $\boldsymbol{\beta}$ is a $(k+1) \times 1$ vector of unknown constants (parameters), and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of unobservable random variables with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$, where $\sigma^2 > 0$ is an unknown constant parameter.

1. For the general linear regression model, what are $E(\mathbf{Y})$ and $cov(\mathbf{Y})$?

2. For the general linear regression model,

   (a) Show (there is no difference beween "show" and "prove") that the matrix $\mathbf{X}'\mathbf{X}$ is symmetric.

   (b) Recall that the $p \times p$ matrix $\mathbf{A}$ is said to be *non-negative definite* if $\mathbf{v}'\mathbf{A}\mathbf{v} \geq 0$ for all constant vectors $\mathbf{v} \in \mathbb{R}^p$. Show that $\mathbf{X}'\mathbf{X}$ is non-negative definite.

   (c) Show that if the columns of $\mathbf{X}$ are linearly independent, then $\mathbf{X}'\mathbf{X}$ is positive definite.

   (d) Show that if $\mathbf{X}'\mathbf{X}$ is positive definite, then $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

   (e) Show that if $(\mathbf{X}'\mathbf{X})^{-1}$ exists, then the columns of $\mathbf{X}$ are linearly independent.

   This is a good problem because it establishes that the least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ exists if and only if the columns of $\mathbf{X}$ are linearly independent, meaning that no independent variable is a linear combination of the other ones.

3. Is $\widehat{\boldsymbol{\beta}}$ an unbiased estimator of $\boldsymbol{\beta}$? Answer Yes or No and show your work.

4. Calculate $cov(\widehat{\boldsymbol{\beta}})$ and simplify. Show your work.

5. Define $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{Y}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Define $\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}}$.

   (a) What are the dimensions of the matrix $\mathbf{H}$?

   (b) Show that $\mathbf{H}$ is symmetric.

   (c) Show that $\mathbf{H}$ is idempotent, meaning $\mathbf{H} = \mathbf{H}^2$

   (d) Show that $\widehat{\boldsymbol{\epsilon}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$.

   (e) Show that $\mathbf{I} - \mathbf{H}$ is symmetric.

   (f) Show that $\mathbf{I} - \mathbf{H}$ is idempotent

   (g) What are the dimensions of the matrix $\widehat{\boldsymbol{\beta}}$?

[1] Copyright information is at the end of the last page.

(h) What are the dimensions of the matrix $\widehat{\mathbf{Y}}$?

(i) What is $E(\widehat{\mathbf{Y}})$? Show your work.

(j) What is $cov(\widehat{\mathbf{Y}})$? Show your work. It is easier if you use $\mathbf{H}$.

(k) What are the dimensions of the matrix $\widehat{\boldsymbol{\epsilon}}$?

(l) What is $E(\widehat{\boldsymbol{\epsilon}})$? Show your work. Is $\widehat{\boldsymbol{\epsilon}}$ an unbiased estimator of $\boldsymbol{\epsilon}$? This is a trick question, and requires thought.

(m) What is $cov(\widehat{\boldsymbol{\epsilon}})$? Show your work. It is easier if you use $\mathbf{I} - \mathbf{H}$.

6. Show that $\mathbf{X}'\hat{\boldsymbol{\epsilon}} = \mathbf{0}$. If the statement is false (not true in general), explain why it is false.

7. The scalar form of the general linear regression model is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i,$$

where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a distribution with expected value zero and variance σ^2. The numbers x_{ij} are known, observed constants, while $\beta_0, \ldots, \beta_k$ and σ^2 are unknown constants (parameters). The term "ranom sample" means independent and identically distributed in this course, so the ϵ_i random variables have zero covariance with one another.

(a) What is $E(Y_i)$?

(b) What is $Var(Y_i)$?

(c) What is $Cov(Y_i, Y_j)$ for $i \neq j$?

8. Starting with the scalar form of the linear regression model (see Question 7), we obtain least-squares estimates of the β values by minimizing the sum of squared differences between observed Y_i and $E(Y_i)$. That is, we choose $\beta_0, \ldots, \beta_k$ to make

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$$

as small as possible.

(a) Differentiate $Q(\boldsymbol{\beta})$ with respect to β_0 and set the derivative to zero, obtaining the first *normal equation*.

(b) Noting that the quantities $\widehat{\beta}_0, \ldots, \widehat{\beta}_k$ (whatever they are) must satisfy the first normal equation, show that the least squares plane must pass through the point $(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k, \overline{Y})$.

(c) Defining "predicted" Y_i as $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_k x_{ik}$, show that $\sum_{i=1}^n \widehat{Y}_i = \sum_{i=1}^n Y_i$.

(d) The *residual* for observation i is defined by $\widehat{\epsilon}_i = Y_i - \widehat{Y}_i$. Show that the sum of residuals equals exactly zero.

9. "Simple" regression is just regression with a single independent variable. The model equation is $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Fitting this simple regression problem into the matrix framework of the general linear regression model,

   (a) What is the $\mathbf{X}$ matrix?

   (b) What is $\mathbf{X}'\mathbf{X}$?

   (c) What is $\mathbf{X}'\mathbf{Y}$?

   (d) What is $(\mathbf{X}'\mathbf{X})^{-1}$?

10. In Question 9, the model had both an intercept and one independent variable. But suppose the model has no intercept. This is called simple *regression through the origin*. The model would be $Y_i = \beta_1 x_i + \epsilon_i$.

    (a) What is the $\mathbf{X}$ matrix?

    (b) What is $\mathbf{X}'\mathbf{X}$?

    (c) What is $\mathbf{X}'\mathbf{Y}$?

    (d) What is $(\mathbf{X}'\mathbf{X})^{-1}$?

11. There can even be a regression model with an intercept but no independent variable. In this case the model would be $Y_i = \beta_0 + \epsilon_i$.

    (a) Find the least squares estimator $\widehat{\beta}_0$ with calculus.

    (b) What is the $\mathbf{X}$ matrix?

    (c) What is $\mathbf{X}'\mathbf{X}$?

    (d) What is $\mathbf{X}'\mathbf{Y}$?

    (e) What is $(\mathbf{X}'\mathbf{X})^{-1}$?

    (f) Verify that your expression for $\widehat{\beta}_0$ agrees with $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

12. Referring to the matrix version of the linear model and letting $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ (which implies that the columns of $\mathbf{X}$ must be linearly independent), show that $(\mathbf{Y} - \widehat{\mathbf{Y}})'(\widehat{\mathbf{Y}} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$.

13. Using the result of the preceding question and writing $Q(\boldsymbol{\beta})$ as $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, show that $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Why does this imply that the minimum of $Q(\boldsymbol{\beta})$ occurs at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$? How do you know that the minimum is unique?

14. The set of vectors $\mathcal{V} = \{\mathbf{v} = \mathbf{Xb} : \mathbf{b} \in \mathbb{R}^{k+1}\}$ is the subset of $\mathbb{R}^n$ consisting of linear combinations of the columns of $\mathbf{X}$. That is, $\mathcal{V}$ is the space *spanned* by the columns of $\mathbf{X}$. The least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ was obtained by minimizing $(\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb})$ over all $\mathbf{b} \in \mathbb{R}^{k+1}$. Thus, $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ is the point in $\mathcal{V}$ that is *closest* to the data vector $\mathbf{Y}$. Geometrically, $\widehat{\mathbf{Y}}$ is the *projection* (shadow) of $\mathbf{Y}$ onto $\mathcal{V}$. The hat matrix $\mathbf{H}$ is a *projection matrix*. It projects the image on any point in $\mathbb{R}^n$ onto $\mathcal{V}$. Now we will test out several consequences of this idea.

    (a) The shadow of a point already in $\mathcal{V}$ should be right at the point itself. Show that if $\mathbf{v} \in \mathcal{V}$, then $\mathbf{Hv} = \mathbf{v}$.

    (b) The vector of differences $\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}}$ should be perpendicular (at right angles) to each and every basis vector of $\mathcal{V}$. How is this related to Question 6?

    (c) Show that the vector of residuals $\widehat{\boldsymbol{\epsilon}}$ is perpendicular to any $\mathbf{v} \in \mathcal{V}$.
