STA 302f16 Assignment Four[1]

The general linear regression model is $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where 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 $cov(\boldsymbol{\epsilon}) = \sigma^2 I_n$. The error variance $\sigma^2 > 0$ is an unknown constant parameter.

1. For the general linear regression model,
   (a) Show (there is no difference beween "show" and "prove") that the matrix $X'X$ is symmetric.
   (b) Show that $X'X$ is non-negative definite.
   (c) Show that if the columns of X are linearly independent, then $X'X$ is positive definite.
   (d) Show that if $X'X$ is positive definite, then $(X'X)^{-1}$ exists.
   (e) Show that if $(X'X)^{-1}$ exists, then the columns of X are linearly independent.

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

2. Let $\widehat{\mathbf{y}} = X\mathbf{b} = H\mathbf{y}$, where $H = X(X'X)^{-1}X'$. The residuals are in the vector $\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$.
   (a) What are the dimensions of the matrix H? Give the number of rows and the number of columns.
   (b) Show that H is symmetric.
   (c) Show that H is idempotent, meaning $H = H^2$
   (d) Using $tr(AB) = tr(BA)$, find $tr(H)$.
   (e) Show that $\mathbf{e} = (I - H)\mathbf{y}$.
   (f) Show that $M = I - H$ is symmetric.
   (g) Show that M is idempotent.
   (h) Using $tr(AB) = tr(BA)$, find $tr(M)$.

3. Please read Chapter 2, pages 28-37 in the textbook.
   (a) This question starts with something you have already done. For the case of simple regression with $k = 1$ independent variables, partially differentiate $\mathcal{S}$ — defined in the first line of (2.6) – with respect the β_0 and β_1. Set both derivatives to zero, obtaining two equations in two unknowns. Now here's the new part. Write these equations in matrix form, obtaining a special case of (2.8).

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

i. What is the $X'X$ matrix? It is a 2×2 matrix with a formula in each cell.

ii. What is the $X'\mathbf{y}$ matrix? It is a 2×1 matrix with a formula in each cell.

(b) Show that $M\boldsymbol{\epsilon} = \mathbf{e}$.

(c) Prove that $X'\mathbf{e} = \mathbf{0}$. If the statement is false (not true in general), explain why it is false.

(d) Prove Theorem 2.1 in the text. I know this is a bit redundant.

(e) Why does $X'\mathbf{e} = \mathbf{0}$ tell you that if a regression model has an intercept, the residuals must add up to zero?

(f) Letting $\mathcal{S} = (\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta})$, show that

$$\mathcal{S} = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) + (\mathbf{b} - \boldsymbol{\beta})'(X'X)(\mathbf{b} - \boldsymbol{\beta}).$$

Why does this imply that the minimum of $\mathcal{S}(\boldsymbol{\beta})$ occurs at $\boldsymbol{\beta} = \mathbf{b}$? The columns of X are linearly independent. Why does linear independence guarantee that the minimum is unique?

(g) What are the dimensions of the random vector $\mathbf{b}$ as defined in Expression (2.9)?

(h) Is $\mathbf{b}$ an unbiased estimator of $\boldsymbol{\beta}$? Answer Yes or No and show your work.

(i) Calculate $cov(\mathbf{b})$ and simplify. Show your work.

(j) What are the dimensions of the random vector $\widehat{\mathbf{y}}$?

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

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

(m) What are the dimensions of the random vector $\mathbf{e}$?

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

(o) What is $cov(\mathbf{e})$? Show your work. It is easier if you use $I - H$.

(p) Prove $E(\mathbf{e}'\mathbf{e}) = \sigma^2(n - k - 1)$

(q) Do Exercises 2.1, 2.3 and 2.6 in the text.

4. 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 $i \neq j$?

5. In *simple regression through the origin*, there is one independent variable and no intercept. The model is $y_i = \beta_1 x_i + \epsilon_i$.

   (a) What is the X matrix?

   (b) What is $X'X$?

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

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

   (e) What is $b_1 = (X'X)^{-1}X'\mathbf{y}$? Compare your answer to (1.22) on page 11 in the textbook.

6. There can even be a regression model with an intercept and no independent variables. In this case the model would be $y_i = \beta_0 + \epsilon_i$.

   (a) Find the least squares estimator of β_0 with calculus.

   (b) What is the X matrix?

   (c) What is $X'X$?

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

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

   (f) What is $b_0 = (X'X)^{-1}X'\mathbf{y}$? Compare this with your answer to Question 6a.

7. The set of vectors $\mathcal{V} = \{\mathbf{v} = X\mathbf{a} : \mathbf{a} \in \mathbb{R}^{k+1}\}$ is the subset of $\mathbb{R}^n$ consisting of linear combinations of the columns of X. That is, $\mathcal{V}$ is the space *spanned* by the columns of X. The least squares estimator $\mathbf{b} = (X'X)^{-1}X'\mathbf{y}$ was obtained by minimizing $(\mathbf{y} - X\mathbf{a})'(\mathbf{y} - X\mathbf{a})$ over all $\mathbf{a} \in \mathbb{R}^{k+1}$. Thus, $\widehat{\mathbf{y}} = X\mathbf{b}$ 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 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 $H\mathbf{v} = \mathbf{v}$.

   (b) The vector of differences $\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$ should be perpendicular (at right angles) to each and every basis vector of $\mathcal{V}$. How is this related to Question 3c?

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