Answers to the 5th problem set

Suppose $\varepsilon_1, \ldots, \varepsilon_n \sim$ i. i. d. $N(0, \sigma^2)$, and for $i = 1, \ldots, n$ we have $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. (Capital "$Y$" and lower-case "$x$"; the former is random; the latter is constant.) This model can be written as

$$Y = X\beta + \varepsilon, \qquad \varepsilon \sim N_n(0,\, \sigma^2 I_n)$$

or as

$$Y \sim N_n(X\beta,\, \sigma^2 I_n).$$

1. Find the entries in the matrix X.

   <u>Answer</u>: "$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i = 1, \ldots, n$" is equivalent to

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \qquad \text{So } X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

2. Let $H = X(X'X)^{-1}X'$. Show that if u is in the column space of X, then $Hu = u$, and if u is orthogonal (i.e., at right angles) to the column space of X, then $Hu = 0$.

   <u>Answer</u>: If u is in the column space of X then for some vector a we have $u = Xa$. Therefore

$$Hu = HXa = \left\{ X(X'X)^{-1}X' \right\} \left\{ Xa \right\} = X \left\{ (X'X)^{-1}(X'X) \right\} a = Xa = u.$$

   "u is orthogonal to the column space" means u is orthogonal to every member of the column space, and that is true if and only if u is orthogonal to every column. If the columns of X are called X_1 and X_2, then u is orthogonal to X_1 iff the dot-product of u with X_1 is 0, and that is the same as $u'X_1 = 0$. Similarly $u'X_2 = 0$. So $u'X = u' \begin{bmatrix} X_1, & X_2 \end{bmatrix} = \begin{bmatrix} 0, & 0 \end{bmatrix}$, or, understanding "0" to mean this two-component row vector, we have $u'X = 0$. That is true iff $X'u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, or, if we understand "0" to be this two-component column vector, $X'u = 0$. So we have

$$Hu = \left\{ X(X'X)^{-1}X' \right\} u = X(X'X)^{-1} \left\{ X'u \right\} = X(X'X)^{-1} \left\{ 0 \right\} = 0.$$

3. Fill in the blanks:
$$(X'X)^{-1}X'Y \sim N_?(?,\ ?).$$
What is $(X'X)^{-1}X'Y$ an unbiased estimator of? (X and Y are "observable.")

Answer: If A is a constant (i.e., non-random) $m \times n$ matrix then $\mathbf{E}(AY) = A\,\mathbf{E}(Y)$ and $\mathbf{var}(AY) = A(\mathbf{var}(Y))A'$. Therefore

$$\begin{aligned} \mathbf{E}((X'X)^{-1}X'Y) &= (X'X)^{-1}X'\,\mathbf{E}(Y) \\ &= (X'X)^{-1}X'X\beta \\ &= \beta, \end{aligned}$$

$$\begin{aligned} \text{and } \mathbf{var}((X'X)^{-1}X'Y) &= (X'X)^{-1}X'(\mathbf{var}(Y))X(X'X)^{-1} \\ &= (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned}$$

So $(X'X)^{-1}X'Y$ is an unbiased estimator of $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$.

4. Use the result of #3 to find an unbiased estimator $\widehat{\beta}_1$ of the slope β_1. Write this estimator in the form
$$\widehat{\beta}_1 = [\text{some row vector (specify and simplify!)}]\,Y.$$
Fill in the blanks:
$$\widehat{\beta}_1 \sim N_?(?,\ ?).$$

Answer: We have

$$\widehat{\beta} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}, \text{ so } \widehat{\beta}_1 = [0,\ 1] \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = [0,\ 1]\widehat{\beta} = [0,\ 1](X'X)^{-1}X'Y.$$

In order to simplify this we first recall two identities:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

and
$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2.$$

CONTINUED⟶

Now

$$
\begin{aligned}
\widehat{\beta}_1 = [0,\ 1](X'X)^{-1}X'Y &= [0,\ 1]\begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} X'Y \\
&= [0,\ 1]\frac{1}{n\left(\left(\sum_{i=1}^n x_i^2\right) - \overline{x}^2\right)}\begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix} X'Y \\
&= \frac{[-\overline{x},\ 1]}{\sum_{i=1}^n (x_i - \overline{x})^2} X'Y \\
&= \frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2}[-\overline{x},\ 1]\begin{bmatrix} 1, & \cdots\cdots, & 1 \\ x_1, & \cdots\cdots, & x_n \end{bmatrix} Y \\
&= \frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2}\left[x_1 - \overline{x}, \ldots\ldots, x_n - \overline{x}\right] Y \\
&= \left[\ \frac{x_1 - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2},\ \ \ldots\ldots,\ \ \frac{x_n - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}\ \right]\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}.
\end{aligned}
$$

Once again we will use the identities $\mathbf{E}(AY) = A\,\mathbf{E}(Y)$ and $\mathbf{var}(AY) = A(\mathbf{var}(Y))A'$:

$$\begin{aligned}
\mathbf{E}\left(\widehat{\beta}_1\right) &= \mathbf{E}\left([0,\ 1]\widehat{\beta}\right) = [0,\ 1]\,\mathbf{E}\left(\widehat{\beta}\right) = [0,\ 1]\beta = \beta_1. \\
\mathbf{var}\left(\widehat{\beta}_1\right) &= [0,\ 1]\left(\mathbf{var}\left(\widehat{\beta}\right)\right)\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= [0,\ 1]\frac{\sigma^2}{n\left(\sum_{i=1}^n (x_i - \overline{x})^2\right)}\begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}.
\end{aligned}$$

$$\text{So } \widehat{\beta}_1 \sim N_1\left(\beta_1,\ \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right).$$

5. Use the result of #4 to find a 90% confidence interval for the slope β_1, assuming (unrealistically) that σ is known. I.e., we want

$$\mathbf{Pr}(\text{some statistic} < \beta_1 < \text{some statistic}) = 0.9.$$

<u>**Answer:**</u> From $\widehat{\beta}_1 \sim N_1\left(\beta_1,\ \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right)$, it follows that

$$\mathbf{Pr}\left(-1.645 < \frac{\widehat{\beta}_1 - \beta_1}{\left(\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right)} < 1.645\right) = 0.9,$$

and thence, that

$$\mathbf{Pr}\left(\widehat{\beta}_1 - 1.645\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} < \beta_1 < \widehat{\beta}_1 + 1.645\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right) = 0.9.$$

6. First some notation:

$$\begin{aligned}
\varepsilon &= Y - X\beta = \text{(unobservable) vector of "errors,"} \\
\widehat{\varepsilon} &= Y - X\widehat{\beta} = \text{(observable) vector of "residuals."}
\end{aligned}$$

Show that $\widehat{\varepsilon} = (I - H)Y$. Use the result, and the identity

$$\mathbf{cov}(AU, BV) = A(\mathbf{cov}(U, V))B' \qquad (A, B \text{ are constant})$$

to find $\mathbf{cov}(\widehat{\varepsilon}, \widehat{\beta}_1)$. From the result, and the joint normality of these two random variables, draw a conclusion about the nature of the dependence between them.

Answer: First observe that $HX = X$. This follows instantly from the definition of H: We have $HX = \left\{X(X'X)^{-1}X'\right\}X = X\left\{(X'X)^{-1}X'X\right\} = X$. From this it follows that $(I-H)X = X - X = 0$. And then:

$$
\begin{aligned}
\mathbf{cov}(\widehat{\varepsilon}, \widehat{\beta}_1) &= \mathbf{cov}\left((I-H)Y,\ \ [0,\ 1](X'X)^{-1}X'Y\right) \\
&= (I-H)\left(\mathbf{cov}(Y,\ Y)\right)\left([0,\ 1](X'X)^{-1}X'\right)' \\
&= (I-H)(\sigma^2 I_n)X(X'X)^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \sigma^2 \underbrace{(I-H)X}_{\substack{\uparrow \\ \boxed{\text{This}=0.}}} (X'X)^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= 0.
\end{aligned}
$$

The answer to #6 continues on the next page.⟶

If $Y \in \mathbb{R}^n$ has a multivariate normal distribution and $C \in \mathbb{R}^{m\times n}$ then CY has a multivariate normal distribution. In particular if $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{1\times n}$, so that $\begin{bmatrix} A \\ B \end{bmatrix} \in \mathbb{R}^{(n+1)\times n}$, then $\begin{bmatrix} A \\ B \end{bmatrix} Y = \begin{bmatrix} AY \\ BY \end{bmatrix}$ has a multivariate normal distribution. In other words, the two random vectors AY and BY are jointly normal. If they are jointly normal and uncorrelated, then they are independent. Therefore $\widehat{\varepsilon}$ and $\widehat{\beta}_1$ are independent.

7. From #2 we conclude that H represents the orthogonal projection onto the column space of X. Use that fact and some facts discussed in class to find the distribution of

$$\frac{\|\widehat{\varepsilon}\|^2}{\sigma^2}$$

and to show this is independent of $\widehat{\beta}_1$.

Answer: Their independence follows immediately from the answer to #6.

Observe that $I - H$ is idempotent:

$$(I-H)(I-H) = I - H - H + H^2 = I - H - H + H = I - H.$$

Since I and H are symmetric, $I - H$ is symmetric. Since $u \mapsto Hu$ is the orthogonal projection onto the 2-dimensional column space of X, it must be that $u \mapsto (I-H)u$ is the orthogonal projection onto the $(n-2)$-dimensional orthogonal complement of the column space of X. Proceeding as in the "attachment" to the summary of March 13th (see ⟨**http://web.mit.edu/18.441/summaries.html**⟩) we conclude that there must be an orthogonal matrix $G = [g_1, \ldots, g_n]$ such that

$$I - H = G \begin{bmatrix} 0 & & \\ & 0 & \\ & & I_{n-2} \end{bmatrix} G' \qquad \text{and} \qquad G'(I-H)G = \begin{bmatrix} 0 & & \\ & 0 & \\ & & I_{n-2} \end{bmatrix}.$$

Multiplication by this diagonal matrix takes $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ to 0, and takes $\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ to 0. Therefore

$$0 = G'(I-H)G \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = G'(I-H)g_1,$$

and since G' is invertible, we get $(I-H)g_1 = 0$. In the same way, we get $(I-H)g_2 = 0$. So g_1 and g_2 must be in the 2-dimensional column space of X, and so $g_3, \ldots, g_n$ must be orthogonal to the column space of X. CONTINUED ⟶

Since multiplication by the diagonal matrix takes $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ to itself, the same sort of reasoning as above shows that $(I-H)g_3 = g_3, \ldots\ldots, (I-H)g_n = g_n$.

Now apply this to $\|\widehat{\varepsilon}\|^2$:

$$\begin{aligned}
\|\widehat{\varepsilon}\|^2 &= \widehat{\varepsilon}'\widehat{\varepsilon} = ((I-H)Y)'((I-H)Y) = Y'(I-H)'(I-H)Y = Y'(I-H)(I-H)Y \\
&= Y'(I-H)Y = Y'G \begin{bmatrix} 0 & & \\ & 0 & \\ & & I_{n-2} \end{bmatrix} G'Y = (G'Y)' \begin{bmatrix} 0 & & \\ & 0 & \\ & & I_{n-2} \end{bmatrix} (G'Y) \\
&= Z' \begin{bmatrix} 0 & & \\ & 0 & \\ & & I_{n-2} \end{bmatrix} Z = Z_3^2 + \cdots\cdots + Z_n^2.
\end{aligned}$$

So we need the distribution of Z:

$$\begin{aligned}
Z = G'Y \ &\sim \ N_n(G' \,\mathbf{E}(Y),\ G'(\mathbf{var}(Y))G) = N_n(G'X\beta,\ G'(\sigma^2 I_n)G) \\
&= \ N_n(G'X\beta,\ \sigma^2 G'G) = N_n(G'X\beta,\ \sigma^2 I_n).
\end{aligned}$$

What do we do with $G'X\beta$? Notice that

$$G'X\beta = \begin{bmatrix} g_1' \\ g_2' \\ g_3' \\ \vdots \\ g_n' \end{bmatrix} X\beta = \begin{bmatrix} g_1'X \\ g_2'X \\ 0 \quad 0 \\ \vdots \quad \vdots \\ 0 \quad 0 \end{bmatrix} \beta = \begin{bmatrix} g_1'X\beta \\ g_2'X\beta \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

since $g_3, \ldots, g_n$ are orthogonal to the column space of X.
Consequently $\mathbf{E}(Z_3) = \cdots = \mathbf{E}(Z_n) = 0$. We have $Z_3, \ldots, Z_n \sim$ i. i. d. $N(0, \sigma^2)$, and so $Z_3^2 + \cdots + Z_n^2 \sim \sigma^2 \chi^2_{n-2}$. Since $\|\widehat{\varepsilon}\|^2 = Z_3^2 + \cdots + Z_n^2$, we have

$$\frac{\|\widehat{\varepsilon}\|^2}{\sigma^2} \sim \chi^2_{n-2}.$$

8. Use the results of (4), (6), and (7) to find a 90% confidence interval for β_1, i.e., two statistics satisfying

$$\mathbf{Pr}(\text{statistic} < \beta_1 < \text{statistic}) = 0.9.$$

Answer: We have found that

$$\widehat{\beta}_1 \sim N_1\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right),$$

$$\|\widehat{\varepsilon}\|^2 \sim \sigma^2 \chi^2_{n-2},$$

and these are independent. Therefore

$$\frac{\widehat{\beta}_1 - \beta}{\left(\dfrac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right)} \sim N(0, 1),$$

$$\frac{\|\widehat{\varepsilon}\|^2}{\sigma^2} \sim \chi^2_{n-2},$$

and these are independent. We know that

$$\frac{N(0,1)}{\sqrt{\chi^2_{n-2}/(n-2)}} = t_{n-2}$$

if the numerator and denominator are independent. So

$$\frac{\left(\widehat{\beta}_1 - \beta_1\right)\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}{\|\widehat{\varepsilon}\|/\sqrt{n-2}} \sim t_{n-2},$$

and "σ" has canceled out. So

$$\mathbf{Pr}\left(-c < \frac{\left(\widehat{\beta}_1 - \beta_1\right)\sqrt{(n-2)\sum_{i=1}^m (x_i - \overline{x})^2}}{\|\widehat{\varepsilon}\|} < c\right) = 0.9$$

where c is the 95th (not 90th!) percentile of Student's distribution with $n-2$ degrees of freedom. If you knew the value of n, you could find the value of c in the table on pages 776-777 of DeGroot & Schervish.

$$\mathbf{Pr}\left(\widehat{\beta}_1 - c\frac{\|\widehat{\varepsilon}\|}{\sqrt{(n-2)\sum_{i=1}^m (x_i - \overline{x})^2}} < \beta_1 < \widehat{\beta}_1 + c\frac{\|\widehat{\varepsilon}\|}{\sqrt{(n-2)\sum_{i=1}^m (x_i - \overline{x})^2}}\right) = 0.9.$$

The leftmost and rightmost members of the inequality are the bounds of the confidence interval.