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.

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 $Hu = 0$. A HINT is in a footnote[1].

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.")

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:
$$\widehat{\beta}_1 \sim N_?(?,\, ?).$$

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

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
$$\operatorname{cov}(AU, BV) = A(\operatorname{cov}(U, V))B' \qquad (A, B \text{ are constant})$$
to find $\operatorname{cov}(\widehat{\varepsilon},\, \widehat{\beta}_1)$. From the result, and the joint normality of these two random variables, draw a conclusion about the nature of the dependence between them.

CONTINUED⟶

[1] A vector u is in the column space of X if and only if $u = Xa$ for some vector a.

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 $\widehat{\beta}_1$.

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

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