7th problem set,

I had not expected to do linear regression very much this early in the course. I assigned a problem on it because of our treatment of confidence intervals, and now I find we're on a roll. #1 and #2 below are on regression. Later we'll see some linear regression problems that are perhaps more concrete than these.

1. Suppose $\varepsilon_1, \dots, \varepsilon_n \sim \text{i.i.d.}\, N(0, \sigma^2)$, and for $i = 1, \dots, n$ we have $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, or, in other words, $Y \sim N_n(X\beta,\, \sigma^2 I_n)$ where
$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$
On the 5th problem set you found that $\widehat{\beta}_1 = [0,\ 1](X'X)^{-1}X'Y = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}$ is an unbiased estimator of β_1. Then, in #2(c) on the 6th problem set you saw this referred to as "the least-squares estimator of β_1." More generally, $\widehat{\beta} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}$ is called "the least-squares estimator of $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$." That means $\widehat{\beta}$ is the value of β that minimizes the sum of squares $\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$. You will justify the nomenclature by showing that $\widehat{\beta} = (X'X)^{-1}X'Y$ does minimize that sum, via the steps outlined below. As before, let $H = X(X'X)^{-1}X'$.
   (a) Show that for any $a, b \in \mathbb{R}^n$ we have $\|a+b\|^2 = \|a\|^2 + 2a'b + \|b\|^2$. (Give a very short answer; don't prove rules of vector or matrix algebra from scratch.)
   (b) Observe that $Y - X\beta = (I - H)Y + (HY - X\beta)$. Then let $a = (I - H)Y$ and $b = HY - X\beta$, and apply the result of (a) and simplify. Finally, show that the sum is less when $\beta = \widehat{\beta}$ than when $\beta =$ anything else.

2. Consider what was done in #2 on the 6th problem set. Weaken the assumptions, as follows: Do not assume $\varepsilon_1, \dots\dots, \varepsilon_n$ are normally distributed, nor that they are identically distributed, nor that they are independent, but assume $\mathbf{E}(\varepsilon) = 0 \in \mathbb{R}^n$ and $\mathbf{var}(\varepsilon) = \sigma^2 I_n$. (Although the errors are not assumed to be identically distributed, they still all have expectations equal to zero and they all have the same variance. Equality of variances is expressed by saying the errors are *homoscedastic* (sometimes spelled *homoskedastic*). And the assumption of independence has been weakened to uncorrelatedness.) As in that problem, assume $Y = X\beta + \varepsilon$, and that X is the same known $n \times 2$ matrix and β is an unknown 2×1 column vector.

   Show that we can still prove the same conclusion by the same method as in #2 on the 6th problem set, i.e., that the least-squares estimator of β is the best linear unbiased estimator of β. (That conclusion under the present assumptions is the Gauss-Markov theorem.)

CONTINUED⟶

3. In families with two or more children, let μ and ν be respectively the average scores of first-born and second-born children on an aptitude test. Suppose these are normally distributed with variance σ^2. A researcher is interested in the difference between the average scores of first- and second-born children, i.e., in $\mu - \nu$.

   WARNING: Be sure you understand the DIFFERENCE between parts (a) and (b) below.

   A HINT is at ⟨**http://web.mit.edu/18.441/assignments.html**⟩.

   (a) Let $X_1, \ldots, X_{12}$ be the scores of first-born children from 12 families chosen independently of each other. Let $Y_1, \ldots, Y_{12}$ be the scores of second-born children from 12 families chosen independently of each other and independently of the first 12 families. Thus we have

$$\begin{aligned} X_1, \ldots, X_{12} &\sim \text{i. i. d.}\, N(\mu, \sigma^2), \\ \text{and } Y_1, \ldots, Y_{12} &\sim \text{i. i. d.}\, N(\nu, \sigma^2), \end{aligned}$$

   and $(X_1, \ldots, X_{12})$ is independent of $(Y_1, \ldots, Y_{12})$. Let $S_1^2 = \sum_{i=1}^{12} \left(X_i - \overline{X} \right)^2$ and let $S_2 = \sum_{i=1}^{12} \left(Y_i - \overline{Y} \right)^2$. Find a pivotal quantity (i.e., a random variable, not necessarily a statistic, whose probability distribution does not depend on the unobservables μ, ν, and σ) whose value depends on $\overline{X} - \overline{Y}$, S_1^2, S_2^2, and $\mu - \nu$, but not on σ, that has a t-distribution and is suitable for finding a 90% confidence interval for $\mu - \nu$. Then find the confidence interval.

   (b) Modify the problem: Instead of picking the second-born children independently of the first-born children, just pick 12 families independently of each other, and use the first- and second-born children from all 12 families. Find a 90% confidence interval for $\mu - \nu$ under these assumptions.

   (c) According to your answers to (a) and (b), which of these two ways of designing the experiment results in a shorter 90% confidence interval for $\mu - \nu$, given the same 24 test scores?

   (d) Explain how the answer to (c) could have been anticipated without doing (a) or (b).

4. (a) DeGroot & Schervish p. 461 #4.

   (b) Modify part (a): Drop the assumption that $n = 25$. Suppose we want

$$\begin{aligned} \mathbf{Pr}(\text{reject } H_0 \mid H_0) &\leq 0.02, \\ \text{and } \mathbf{Pr}(\text{reject } H_0 \mid \mu = \mu_0 + 0.3) &\geq 0.8. \end{aligned}$$

   In other words, the probability of Type I error, i.e., the probability of rejecting the null hypothesis given that the null hypothesis is true, is no more than 2%, and if the null hypothesis is false, so that $\mu = \mu_0 + 0.3$ instead of $\mu = \mu_0$, then we have at least an 80% chance of correctly rejecting the null hypothesis. In other words, the probability of Type II error — of failing to reject the null hypothesis in this case — is no more than 20%.

   How big a sample do we need to achieve these desiderata, i.e., how big must n be?

#4 CONTINUES⟶

(c) Now reinstate the assumption that $n = 25$, but alter the alternative hypothesis, so that we have:

$$\begin{aligned} H_0 &: \quad \mu = \mu_0, \\ H_1 &: \quad \mu > \mu_0. \end{aligned}$$

Redo part (a) under these assumptions.