Midterm

William Matzko

March 25, 2021

Problem 7

I don't recall ever seeing a summation without an argument. I know what you mean, but I think it is more common to put a "1" as an argument. I could be wrong, however.

We first note that

$$\begin{aligned}\sum_i^n (x_i - \overline{x})c &= c\sum_i^n x_i - c\overline{x}\sum_i^n \\ &= cn\overline{x} - c\overline{x}n \\ &= 0\end{aligned}$$

where c is any constant. It follows that b can be written as

$$\begin{aligned}b &= \frac{\sum_i^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_i^n (x_i - \overline{x})^2} \\ &= \frac{\sum_i^n (x_i - \overline{x})y_i}{\sum_i^n (x_i - \overline{x})^2} \\ &= \frac{\sum_i^n (x_i - \overline{x})(\beta x_i + \alpha + \epsilon_i)}{\sum_i^n (x_i - \overline{x})^2} \\ &= \frac{\sum_i^n (x_i - \overline{x})\beta x_i}{\sum_i^n (x_i - \overline{x})^2} + \frac{\sum_i^n (x_i - \overline{x})\epsilon_i}{\sum_i^n (x_i - \overline{x})^2}\end{aligned}$$

The expectation value of this is then

$$\begin{aligned}E[b] &= E\left[\frac{\sum_i^n (x_i - \overline{x})\beta x_i}{\sum_i^n (x_i - \overline{x})^2} + \frac{\sum_i^n (x_i - \overline{x})\epsilon_i}{\sum_i^n (x_i - \overline{x})^2}\right] \\ &= E\left[\frac{\sum_i^n (x_i - \overline{x})\beta x_i}{\sum_i^n (x_i - \overline{x})^2}\right] + E\left[\frac{\sum_i^n (x_i - \overline{x})\epsilon_i}{\sum_i^n (x_i - \overline{x})^2}\right]\end{aligned}$$

we note that the x_i terms are fixed input values. In fact, the only term that varies in the above equation is ϵ_i. Because we are drawing this noise term from a Gaussian with $\mu = 0$, the expectation value is simply

$$E[\epsilon_i] = 0$$

Hence, the second term is zero. The expectation value of a constant is just the constant itself, so we may write

$$\begin{aligned}E[b] &= \frac{\sum_i^n (x_i - \overline{x})\beta x_i}{\sum_i^n (x_i - \overline{x})^2} \\ &= \beta\frac{\sum_i^n (x_i - \overline{x})x_i}{\sum_i^n (x_i - \overline{x})^2}\end{aligned}$$

The fraction ends up being 1. To see this, we note that the denominator can be rewritten as

$$\begin{aligned}
\sum_i^n (x_i - \overline{x})^2 &= \sum_i^n (x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\
&= \sum_i^n x_i^2 - 2\overline{x}\sum_i^n x_i + n\overline{x}^2 \\
&= \sum_i^n x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 \\
&= \sum_i^n x_i^2 - n\overline{x}^2 \\
&= \sum_i^n (x_i^2 - \overline{x}x_i) \\
&= \sum_i^n (x_i - \overline{x})x_i
\end{aligned}$$

Hence, the fraction is one. Thus, we may conclude that

$$E[b] = \beta$$

The expectation value of the sample slope is indeed equal to the population slope.

Problem 8

The plot of slope and y-intercept residuals clearly shows a strong negative linear trend. We see that larger slope residuals (i.e. larger positive errors in the slope estimate) necessarily lead to more negative y-intercept residuals. This negative linear correlation easily demonstrates that the errors in a and b are dependent on each other; if they weren't, we would expect to see no discernible trend when plotting the residuals against each other. Such a trend is expected. If we have a larger positive error on our slope, it is natural to 'compensate' for that by having a smaller, more negative y-intercept.

Good that you came up with a physical explanation!