Unit 4 Homework: Conditional Operators and Best Linear Predictors

w203: Statistics for Data Science

Safety First through Statistics

Suppose the strength of a particular metal beam is given by,

$$S = 10 + .5T^2 \cdot P$$

Where T is a random variable representing the forging temperature and P is a random variable representing purity. Suppose the following statements are true about these random variables:

- T has a uniform distribution on $[0, 2]$.
- Conditional on a value for T, P is has a normal distribution with mean $T/2$ and standard deviation $T/12$. For example, $E[P|T=1] = \frac{1}{2}$, and $\sqrt{V[P|T=1]} = \sigma_{(P|T=1)} = \frac{1}{12}$

Answer the following question:

1. (6 points) Compute the expectation of S.

Reasoning about a BLP

Suppose that discrete random variables X and Y have joint probability mass function given by:

$$f(x,y) = \begin{cases} 1/2, & (x,y) \in \{(0,0),(2,1)\} \\ 0, & \text{otherwise} \end{cases}$$

(This means that there is equal probability that the points $(0, 0)$ and $(2, 1)$ are drawn; there is zero probability that any other point is drawn.)

Let $g(x) = \beta_0 + \beta_1 x$ be a predictor for y, and define the error, $\epsilon = Y - g(X)$.

1. (3 points) If you impose the moment condition, $E[\epsilon] = 0$, what one point in the plane must the predictor pass through? (In some places, this point is referred to as the *grand mean.*)
2. (3 points) Assuming $E[\epsilon] = 0$, write an expression for $cov[X, \epsilon] = cov[X, Y - g(x)]$ in terms of β_1.
3. (3 points) In your own words, explain how the sign of $cov[X, \epsilon]$ is related to the angle of the line.
4. (3 points) What predictor fulfills both $E[\epsilon] = 0$ and $cov[X, \epsilon] = 0$?

Note: Maximum score on any homework is 100%