**Conditional and Marginal Means and Variances**

**Marginal Variance:** The definition of the (population) (marginal) variance of a random variable $Y$ is

$$\text{Var}(Y) = E((Y - E(Y))^2)$$

What does this say in words (and pictures)?

There is another formula for $\text{Var}(Y)$ that is sometimes useful in computing variances or proving things about them. It can be obtained by multiplying out the squared expression in the definition:

$$\text{Var}(Y) = E((Y - E(Y))^2) = E(Y^2 - 2YE(Y) + [E(Y)]^2)$$

(Fill in details, and say the final result in words!)

**Conditional Variance:** Similarly, if we are considering a conditional distribution $Y|X$, we define the conditional variance

$$\text{Var}(Y|X) = \text{Variance of the conditional distribution } Y|X$$

$$= E((Y - E(Y|X))^2 | X)$$

(Note that both expected values here are conditional expected values.)

What does this say in words (and pictures)?

Exercise: Derive another formula for the conditional variance, analogous to the second formula for the marginal variance. (And say it in words!)

**Conditional Variance as a Random Variable:** As with $E(Y|X)$, we can consider $\text{Var}(Y|X)$ as a random variable. For example, if $Y =$ height and $X =$ sex for persons in a certain population, then $\text{Var}(\text{height } | \text{sex})$ is the variable which assigns to each person in the population the variance of height for that person's sex.
**Expected Value of the Conditional Variance:** Since \( \text{Var}(Y|X) \) is a random variable, we can talk about its expected value. Using the formula \( \text{Var}(Y|X) = E(Y^2|X) - [E(Y|X)]^2 \), we have

\[
E(\text{Var}(Y|X)) = E(E(Y^2|X)) - E([E(Y|X)]^2)
\]

We have already seen that the expected value of the conditional expectation of a random variable is the expected value of the original random variable, so applying this to \( Y^2 \) gives

\[ (*) \quad E(\text{Var}(Y|X)) = E(Y^2) - E([E(Y|X)]^2) \]

**Variance of the Conditional Expected Value:** For what comes next, we will need to consider the variance of the conditional expected value. Using the second formula for variance, we have

\[
\text{Var}(E(Y|X)) = E([E(Y|X)]^2) - [E(E(Y|X))]^2
\]

Since \( E(E(Y|X)) = E(Y) \), this gives

\[ (**) \quad \text{Var}(E(Y|X)) = E([E(Y|X)]^2) - [E(Y)]^2. \]

**Putting It Together:**

Note that (*) and (**) both contain the term \( E([E(Y|X)]^2) \), but with opposite signs. So adding them gives:

\[
E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) = E(Y^2) - [E(Y)]^2,
\]

which is just \( \text{Var}(Y) \). In other words,

\[ (***) \quad \text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)). \]

In words: The marginal variance is the sum of the expected value of the conditional variance and the variance of the conditional means.
Consequences:

I) This says that two things contribute to the marginal (overall) variance: the expected value of the conditional variance, and the variance of the conditional means. (See Exercise) Moreover, \( \text{Var}(Y) = \text{E}(\text{Var}(Y|X)) \) if and only if \( \text{Var}(\text{E}(Y|X)) = 0 \). What would this say about \( \text{E}(Y|X) \)?

II) Since variances are always non-negative, (***) implies

\[
\text{Var}(Y) \geq \text{E}(\text{Var}(Y|X)).
\]

III) Since \( \text{Var}(Y|X) \geq 0 \), \( \text{E}(\text{Var}(Y|X)) \) must also be \( \geq 0 \). (Why?). Thus (***) implies

\[
\text{Var}(Y) \geq \text{Var}(\text{E}(Y|X)).
\]

Moreover, \( \text{Var}(Y) = \text{Var}(\text{E}(Y|X)) \) if and only if \( \text{Var}(\text{E}(Y|X)) = 0 \). What would this imply about \( \text{Var}(Y|X) \) and about the relationship between \( Y \) and \( X \)?

IV) Another perspective on (***) (cf. Textbook, pp. 36 - 37):

Note that:

i) \( \text{E}(\text{Var}(Y|X)) \) is a weighted average of \( \text{Var}(Y|X) \)

ii) \( \text{Var}(\text{E}(Y|X)) = \text{E}([\text{E}(Y|X) - \text{E}(\text{E}(Y|X))]^2) = \text{E}([\text{E}(Y|X) - (\text{E}(Y))]^2), \) which is a weighted average of \( [\text{E}(Y|X) - (\text{E}(Y))]^2 \)

Thus, (***) says that \( \text{Var}(Y) \) is a weighted mean of \( \text{Var}(Y|X) \) plus a weighted mean of \( [\text{E}(Y|X) - (\text{E}(Y))]^2 \) (and is a weighted mean of \( \text{Var}(Y|X) \) if and only if all conditional expected values \( \text{E}(Y|X) \) are equal to the marginal expected value \( \text{E}(Y) \).)
EXERCISE: What contributes most to Var(Y): Var(E(Y|X)) or E(Var(Y|X))? 

A. 

B. 

C.