Unit 2 Quiz B

Randomly select 3 questions to ask members of the group requesting the next set of assignments. If they get all questions correct, give them the entire stack of quizzes and assignments so that they can quiz the next group. If they get any question wrong, send them back to their seats to review and then try again in 10 minutes or more.

1. In the equation $y_i = \beta_0 + \beta_1 x_i + u_i$, explain what β_0 and β_1 represent. How are they different from $\hat{\beta}_0$ and $\hat{\beta}_1$?

   Answer: β_0 and β_1 are the true population parameters (intercept and slope) in the underlying data generating process. $\hat{\beta}_0$ and $\hat{\beta}_1$ are our estimates of these parameters based on the sample data.

2. Explain why the statement $\sum_i x_i y_i = \sum_i x_i \sum_i y_i$ is false. Give a simple numerical example to demonstrate.

   Answer: Consider x = {1,2} and y = {2,2}. Then $\sum xy = 6$ but $(\sum x)(\sum y) = (3)(4) = 12$. The sum of products does not equal the product of sums because each product $x_i y_i$ captures the relationship between paired observations.

3. What is the key difference between u_i (disturbance term) and e_i (residual)? Which one must sum to zero and why?

   Answer: u_i is the true unobservable error term in the population model, while e_i is the observable residual from our estimated model. $\sum_i e_i$ must equal zero by construction of OLS, but $\sum_i u_i$ need not equal zero.

4. What does OLS minimize? Write the mathematical expression.

   Answer: OLS minimizes the sum of squared residuals: $\sum_i e_i^2$.

5. If a regression coefficient has a p-value of 0.03, can you reject the null hypothesis at the 5% level? What about at the 1% level?

   Answer: Yes at 5% level since 0.03 < 0.05, but no at 1% level since 0.03 > 0.01. The p-value represents the probability of observing such an extreme coefficient value if the null hypothesis were true.

6. How do sample size (n), variance in x, and variance in u affect the precision of OLS estimates?

   Answer: Larger n reduces SE (denominator), larger Var(x) reduces SE (denominator), larger Var(u) increases SE (numerator). This matches intuition about when estimates should be more precise.