Practice Midterm

EC421 Fall 2022

Colleen O'Briant

October 24, 2022

This is a group assignment: please put everyone's names on the top of the assignment and turn in only one solution before the due date on Friday, October 28th at 5pm. If your group gets the highest score in the class on this practice midterm, I will give each member of the group an extra 5 points of extra credit on their midterm. If your group comes in 2nd or 3rd place, I will give each group member an extra 3 and 2 points respectively. Ties will be dealt with in this order: if there is a 2-way tie for first place, I will give each group member on both teams 3 points of extra credit on their midterms and 2 points to each member of the next highest group.

The midterm exam will be in-class (closed-note, completed individually) on Wednesday, November 2nd. It will have 30 multiple choice questions (each worth 1.5 points), 10 short answer questions (each worth 4 points), and 10 tidyverse questions (each worth 1.5 points). You should make sure to study the classwork, the workbook (chapters 1-6), and the koans: anything we've studied so far is fair game.

1 Multiple Choice

1. Consider this model:

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 sex_i + \beta_3 education * sex_i + u_i$$

   Where education is the number of years of education someone has and sex is 0 for females and 1 for males. Then the expected wage for females with 10 years of education is:

   (a) β_0

(b) $\beta_0 + \beta_2 + 10\ \beta_1 + 10\ \beta_3$

(c) $\beta_0 + \beta_2$

(d) $\beta_0 + 10\ \beta_1$

2. Suppose this is the true data generating process for y:

   $y_i = \beta_0 + \beta_1\ x_i + \beta_2\ z_i + u_i$

   If z_i is omitted, will β_1 always be biased?

   (a) No, β_1 will only be biased if z also correlates with x.

   (b) No, β_1 will only be biased if z also correlates with u.

   (c) No, β_1 will never be biased if z is omitted because β_1 is the coefficient on x_i, not on z_i.

   (d) Yes, β_1 will always be biased because, as long as β_2 is nonzero, z has an important effect on y that must be accounted for.

3. Simplify:

   $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

   (a) $n^2\bar{x}\bar{y} - n\bar{x}\bar{y}$

   (b) $\sum_i x_i y_i$

   (c) 0

   (d) $\sum_i x_i y_i - \bar{x}\bar{y}n$

4. In a simple regression, the standard error for $\hat{\beta}_1$ depends _ on the sample variance of the explanatory variable x.

   (a) Positively

   (b) Negatively

   (c) Not at all

5. (T/F): Weighted least squares (WLS) "up-weights" observations with low-variance disturbances and "down-weights" observations with high-variance disturbances.

   (a) True

   (b) False

6. Under standard OLS assumptions, the key assumption for $\hat{\beta}_1$ to be unbiased is _ and the key assumption for $\hat{\beta}_1$ to be consistent is _.

   (a) $\mathrm{E}[\mathrm{u_i} \mid \mathrm{X}] = 0$; $\mathrm{E}[\mathrm{u}] = 0$
   (b) $\mathrm{E}[\mathrm{u_i} \mid \mathrm{X}] = 0$; $\mathrm{Cov}(\mathrm{x_i}, \mathrm{u_i}) = 0$
   (c) x and u are independent; $\mathrm{Cov}(\mathrm{x_i}, \mathrm{u_i}) = 0$
   (d) x and u are independent; $\mathrm{E}[\mathrm{u}] = 0$

7. (T/F): Selection bias disappears with large sample sizes (over 1 million rows).

   (a) True
   (b) False

8. (T/F): It is always true that $\mathrm{E}[\mathrm{u\ x}] = \mathrm{E}[\mathrm{ux} \mid \mathrm{X}]$

   (a) True
   (b) False

9. If our signficance level is 0.01 and our p-value is 0.02, then we [blank] the null hypothesis.

   (a) Reject
   (b) Fail to reject

10. Cov(X, 5X - Y) is equal to:

   (a) 25 Var(X) + Cov(X, Y)
   (b) 25 Var(X) + Var(Y)
   (c) 5 Var(X) - Cov(X, Y)
   (d) 5 Var(X) - Var(Y)

2 Short Answer

Consider a linear model with the intercept omitted:

$y_i = \beta_0 x_i + u_i$

In the next three problems, we'll build the econometrics of this model from scratch.

1. Recall that OLS minimizes the sum of squared residuals. Write down what that means in this context and take first order conditions. Show that $\hat{\beta}_0 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$.

2. Next, show that as long as the model is correctly specified (assume it reflects the true data generating process), $\hat{\beta}_0 = \beta_0 + \frac{\sum_i x_i u_i}{\sum_i x_i^2}$.

3. Finally, let $w_i = \frac{x_i}{\sum) i x_i^2}$ and show that $\hat{\beta}_0$ is an unbiased estimator of β_0 under one crucial assumption.

Use this information to solve questions 4 and 5: suppose we have data from a survey that asks people the number of children they have and their happiness level (scored from 1-10). We could theoretically use this data to answer the question: "Does having children cause people to be happier?" with this model:

$\text{happiness}_i = \beta_0 + \beta_1 \text{number of children}_i + u_i$

4. Suppose also that the survey failed to ask whether the person was single or married. Consider a variable *single*, which is 1 for single people and 0 for married people. Since single is unobserved, it is absorbed into u_i. What does exogeneity mean in this context with respect to *single*? Is exogeneity a fair assumption here? Explain. (Hint: Two people enter a room and you only know their X value. . .)

5. Would omitting *single* bias the estimate for the effect of children on happiness up, down, or not at all? Explain by mentioning $Cov(x, u)$ and β_2 as we did in the classwork.