

**Microeconometría aplicada**  
**Prof. Mauricio Romero**  
**Taller para estudiar para el parcial 2**

- Pueden usar R, Stata o cualquier otro programa
  - Este taller se parece mucho al examen. Durante el examen, deben responder a las preguntas por medio de “Forms”. El examen se cerrara a las 5:30 PM, y no se admiten respuestas tarde.
  - Este es un examen de libro/Google/WhatsApp/Instagram/Twitter/etc. abierto. Usar recursos adicionales (internet, libros, blogs, etc.) está bien (y muchas veces es necesario), pero por favor menciones los recursos que usan y cítelos adecuadamente. Sin embargo, tengan en cuenta el tiempo. Pueden gastarse todo el examen buscando la respuesta de una sola pregunta si no saben la respuesta desde antes.
  - Es individual
  - Para todas las puede usar simulaciones (mínimo 5,000) o hacer los cálculos exactos.
  - En el examen puede subir PDFs o códigos de R justificando sus respuestas. Solo se admiten reclamos que tengan estos archivos.
1. **Verdadero/Falso** Si el impacto promedio del tratamiento (ATE) es 0.5, esto implica que para todos los individuos  $Y_{1i} > Y_{0i}$
  2. **Verdadero/Falso** Si para todos los individuos  $Y_{1i} > Y_{0i}$ , esto implica que el impacto promedio del tratamiento (ATE) es positivo
  3. **Verdadero/Falso** Si el coeficiente de una regresión es diferente de cero, entonces X tiene un efecto causal en Y
  4. **Verdadero/Falso** Si el p-valor de una regresión es  $< 0.001$ , entonces X tiene un efecto económicamente importante en Y
  5. **Verdadero/Falso** Si la correlación entre los residuos y las Y es cero, entonces los estimadores tienen una interpretación causal
  6. **Verdadero/Falso:** Si la CEF no es lineal entonces la estimación con una regresión lineal no es causal.
  7. Si para un tercio de los individuos  $Y_{1i} - Y_{0i} = -0.5$ , para otro tercio  $Y_{1i} - Y_{0i} = 0$  y para otro tercio  $Y_{1i} - Y_{0i} = 0.2$ . ¿Cuánto es el impacto promedio del tratamiento (ATE)?
  8. Suponga que  $Y_0 \sim N(3, 1)$ . Hay un tratamiento que tiene un efecto  $\delta_i \sim N(0, 2)$ . ¿Cuánto es el impacto promedio del tratamiento (ATE)?

Suponga que tiene una muestra de N individuos y está estudiando el efecto del ingreso en la felicidad. Para eso usa una encuesta del INEGI. Pero es bien sabido que entre mas ingresos, mas probabilidad de que las personas no respondan a una encuesta. En particular, si  $I$  es el percentil del ingreso (que va entre 0-1), entonces la probabilidad de responder la encuesta es una variable binomial con parámetro  $1 - I$  (i.e., la probabilidad de responder es 1 menos el percentil del ingreso) para cada individuo. El modelo verdadero de generación de datos es:

$$Y = 1 + 1.5E + 2 * I + \varepsilon$$

Donde  $E$  es la educación, que se distribuye  $E \sim U(10, 20)$ ,  $I$  es el percentil del ingreso, que por definición se distribuye  $I \sim U(0, 1)$  y  $\varepsilon$  es el error que se distribuye  $\varepsilon \Xi(5)$ .

9. ¿Cuánto es el impacto promedio de la educación en la felicidad (ATE) en la población?
10. ¿Cuánto es el impacto estimado sobre la muestra de la encuesta del INEGI?
11. ¿Cuánto el sesgo causado por la selección?

Ahora suponga que en realidad uno no puede medir el percentil del ingreso. Entonces no puede controlar por este factor en la regresión.

12. ¿Cuánto es el impacto promedio de la educación en la felicidad (ATE) en la población?

13. ¿Cuánto es el impacto estimado sobre la muestra de la encuesta del INEGI?

14. ¿Cuánto el sesgo causado por la selección?

Donde  $E$  es la educación, que se distribuye  $E \sim U(10, 20) * I$  (y que el ingreso sigue sin ser observable).

15. ¿Cuánto es el impacto promedio de la educación en la felicidad (ATE) en la población?

16. ¿Cuánto es el impacto estimado sobre la muestra de la encuesta del INEGI?

17. ¿Cuánto el sesgo causado por la selección?

Finalmente, suponga que la probabilidad de responder la encuesta depende del percentil de la felicidad y que observa el percentil del ingreso.

18. ¿Cuánto es el impacto promedio de la educación en la felicidad (ATE) en la población?

19. ¿Cuánto es el impacto estimado sobre la muestra de la encuesta del INEGI?

20. ¿Cuánto el sesgo causado por la selección?

Suponga que usted está encargado de diseñar un experimento para probar si la vacuna “patria” del COVID-19 que “México está desarrollando”.<sup>1</sup> En ausencia de la vacuna, si una persona tiene un evento adverso o no se distribuye Bernoulli con  $p = 0.9$ . Ud. puede reclutar a 100 personas para el experimento (50 de control y 50 de tratamiento). Use inferencia de aleatorización para las siguientes preguntas:

21. Suponga que el efecto verdadero de la vacuna es  $\beta_i = 0$  para la mitad de la población y  $\beta_i = -0.5$  para la otra mitad (es decir, reduce la probabilidad de un evento adverso en 50 puntos porcentuales para la mitad de la población). ¿Cuál es la probabilidad de rechazar la hipótesis nula de que la vacuna no tiene efecto? [Pista: es una prueba de hipótesis de dos colas]

22. Ahora suponga que el efecto del tratamiento depende de una constante multiplicativa  $Y_i^1/Y_i^0 = C$ . y el efecto verdadero de la vacuna es  $\beta_i = 0$  para toda la población. ¿Cuál es la probabilidad de rechazar la hipótesis nula de que la vacuna no tiene efecto? [Pista: es una prueba de hipótesis de una cola]

23. Ahora suponga que tiene un problema de *outliers* y el efecto verdadero de la vacuna es  $\beta_i = 0$  para toda la población. ¿Cuál es la probabilidad de rechazar la hipótesis nula de que la vacuna no tiene efecto si su estadístico  $S(T, Y)$  es la diferencia de quantiles del percentil 50.

Ahora use mínimos cuadrados ordinarios para las siguientes preguntas (y use un nivel de significancia del 5% para rechazar la hipótesis nula):

24. Suponga que el efecto verdadero de la vacuna es  $\beta_i = 0$  para todo  $i$ . ¿Cuál el valor esperado de  $\widehat{\beta}_{OLS}$ , condicional a ser estadísticamente significativo? [Pista: es una prueba de hipótesis de dos colas]

Suponga que Ud. tiene un proceso de generación de datos  $Y = X\beta + \varepsilon$ , donde  $\varepsilon \sim N(3, 10)$ ,  $X \sim N(5, 10)$ ,  $cov(\varepsilon, X) = 3$  y  $\beta = 1$ . Usted tiene una muestra de tamaño 1,000.

25. Calcule  $\mathbb{E}(\widehat{\beta})$

26. Calcule  $V(\widehat{\beta})$

27. ¿Es  $\widehat{\beta}$  un estimador consistente de  $\beta$ ?

28. Calcule  $\bar{\widehat{\varepsilon}} = \frac{1}{1000} \sum_{i=1}^N \widehat{\varepsilon}_i$

29. Calcule  $Cov(x, \widehat{\varepsilon}) = \frac{1}{1000} \sum_{i=1}^N (x_i - \bar{x})(\widehat{\varepsilon}_i - \bar{\widehat{\varepsilon}})$

<sup>1</sup>Para aprender más sobre esto puede leer el siguiente hilo de Twitter <https://twitter.com/TheSJuan/status/1382176227562168321?s=08> o en el “vaccine tracker” del NYT <https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html>.

30. Calcule  $Cov(\hat{y}, \hat{\varepsilon}) = \frac{1}{1000} \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})$

Suponga que queremos saber cuál es el efecto de estudiar economía en el ingreso. Suponga que el efecto causal de estudiar economía en el ingreso es 50,000 MXN para las personas de preparatoria pública y 30,000 MXN para aquellos de preparatoria privada. El 70 % de la población se gradúa de preparatoria pública y el 30 % de preparatoria privada. Entre aquellos de preparatoria pública el 10 % estudia economía, entre los de preparatoria privada el 30 % estudia economía.

31. ¿Cuánto es el impacto promedio del tratamiento (ATE)?
32. Suponiendo que la regresión es causal después de controlar por el tipo de preparatoria del que se graduó una persona: ¿Cuánto es el parámetro que identifica mínimos cuadrados ordinarios en la población?
33. Suponiendo que primero hace una regresión de si se graduado de economía o no contra el tipo de preparatoria y guarda los residuos. Después hace una regresión del ingreso contra esos residuos. ¿Cuánto es el parámetro que identifica esta regresión en la población?

Suponga que queremos saber cuál es el efecto de privatizar escuelas públicas. Suponga el proceso de generación de datos es tal que  $Y_{is} = \alpha + T_s\beta + \varepsilon_s + \varepsilon_i$  donde  $\varepsilon_s$  es un error que es igual para todas las personas de una misma escuela y  $\varepsilon_i$  es un error a nivel individual. Suponga que  $\alpha \sim U[0, 500]$ ,  $\varepsilon_s \sim N(0, 100)$  y  $\varepsilon_i \sim N(0, 150)$ . Usted tiene un experimento con 200 escuelas (mitad de tratamiento y mitad de control). En cada escuela tiene datos de 30 alumnos.

34. Suponga que el efecto verdadero es  $\beta = 0$ . ¿Cuál es la probabilidad de rechazar la hipótesis nula usando errores robustos?
35. Suponga que el efecto verdadero es  $\beta = 10$ . ¿Cuál es la probabilidad de rechazar la hipótesis nula usando errores robustos?
36. Suponga que el efecto verdadero es  $\beta = 0$ . ¿Cuál es la probabilidad de rechazar la hipótesis nula usando errores clúster a nivel escuela?
37. Suponga que el efecto verdadero es  $\beta = 10$ . ¿Cuál es la probabilidad de rechazar la hipótesis nula usando errores clúster a nivel escuela?