

Econometría I

Tarea 3

Magíster en Economía, Universidad Alberto Hurtado

October 8, 2025

1. (20 puntos) Suponga el modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

donde $\mathbb{E}(u_i) = 0$, $\mathbb{E}(x_{1i} u_i) = 0$ y $\mathbb{E}(u_i^2 | x_{1i}) = \sigma^2$ (homoscedasticidad).

- (a) Calcule la varianza asintótica de $\hat{\beta}_1$.

Pista: Hay dos formas de resolver el problema. Puede utilizar la fórmula de la varianza estimada bajo homoscedasticidad con $x_i = (1, x_{1i})'$ o puede utilizar la distribución asintótica de la covarianza derivada en la anterior.

- (b) Explique el efecto sobre la varianza asintótica de β_1 de los siguientes factores: tamaño de la muestra, varianza de x_1 y R^2 .
2. (10 puntos) Un estudio reciente ha encontrado que la tasa de mortalidad de las personas que duermen entre 6 y 7 horas por noche es menor que la tasa de mortalidad de aquellas personas que duermen 8 o más horas. El estudio utilizó 1.1 millón de observaciones de una muestra aleatoria de chilenos entre 30 y 102 años de edad. Cada persona en la muestra fue seguida por 4 años. La tasa de mortalidad para personas que duermen 7 horas fue calculada como el número de muertes durante los 4 años entre las personas que duermen 7 horas sobre el número de personas que duermen 7 horas en la muestra. Cálculos similares se realizaron sobre el resto de las categorías. En base a los resultados del estudio, ¿usted recomendaría a los chilenos que duermen 9 horas por noche reducir las horas de sueño a 6 o 7 horas por noche si quieren prolongar sus vidas? ¿Por qué o por qué no? Explique.
 3. (30 puntos) Utilice los datos de `veteran.dta` del paper de Angrist (Etca, 1998) sobre el efecto de ser veterano de guerra en ingresos.
 - (a) Utilice un modelo de regresión simple para estimar el efecto de ser veterano de guerra (`dvet`) en ingresos (`earnvar`) para blancos y no blancos (`dnwhite`). Por razones de confidencialidad los datos tienen cierta agregación. Para trabajar con este tipo de datos utilice la opción `[pweight = popcount]` en Stata. Interprete los resultados. ¿Usted piensa que el coeficiente estimado mide una relación causal? Explique.

Nota: Realice dos estimaciones distintas, una para blancos y otra para no blancos.

- (b) El ejército de USA realiza un filtro de las aplicaciones en base a un test de habilidad, educación, edad y otras variables observables. Al omitir los datos del test de habilidad y educación, ¿usted piensa que la estimación del efecto ser veterano de guerra en (a) tiene un sesgo positivo o negativo? Explique.
- (c) Estime el efecto de ser veterano de guerra para blancos y no blancos controlando por los distintos valores del test de habilidad (*afqtgrp*) y el nivel de educación (*edgrp*). Interprete la estimación del efecto de ser veterano de guerra, compare con (a) y explique la diferencia.

Nota: Utilice una dummy por cada valor de *afqtgrp* y *edgrp*, pero no por las interacciones.

4. (30 puntos) Utilice los datos de WAGE2.dta del paper de Blackburn and Neumark (QJE, 1992) sobre los retornos de la educación.

- (a) (10 puntos) Utilice un modelo de regresión simple log-nivel para estimar el efecto de un año más de educación (*educ*) en salarios (*wage*). Interprete los resultados. ¿Usted piensa que el coeficiente estimado mide una relación causal? Explique.
- (b) (5 puntos) El modelo de regresión simple omite, entre otras cosas, la habilidad innata. Al omitir los datos de habilidad innata, ¿usted piensa que la estimación del efecto de un año adicional de educación en (a) tiene un sesgo positivo o negativo? Explique.
- (c) (5 puntos) En los datos existe una proxy de habilidad innata: *IQ*. Incluya *IQ* como variable explicativa adicional. Interprete la estimación de un año adicional de educación, compare con (a) y explique la diferencia.
- (d) (5 puntos) La fórmula poblacional del sesgo de variable omitida se cumple en la muestra si reemplazamos momentos poblacionales por momentos muestrales. Primero, compruebe este resultado en la muestra. Regrese *IQ* en *educ* y, en base a este resultado y lo obtenido en (c), recupere la estimación de la pendiente en (a). Segundo, demuestre analíticamente que este resultado siempre se cumple.
- (e) (5 puntos) Estime el efecto de educación sobre salarios controlando por *IQ*, *exper*, *tenure*, *married*, *south* y *urban*. Compare con el resultado de (a) y (c).

5. (30 puntos) Utilice los datos de HPRICE2.dta del paper de Harrison y Rubinfeld (JEEM, 1978) sobre el efecto de la polución en el precio de las viviendas. Los datos son para el área metropolitana de Boston en 1970.

- (a) Utilice un modelo de regresión simple log-log para estimar el efecto de un incremento en la concentración de óxido de nitrógeno en el aire (*nox*) en los precios de las viviendas (*price*). Interprete los resultados. ¿Usted piensa que el coeficiente estimado mide una relación causal? Explique.
- (b) Agregue las siguientes variables explicativas al modelo: número de habitaciones (*rooms*), distancia al centro ($\log(dist)$), y ratio estudiantes por profesor (*stratio*). Interprete la

estimación del efecto de un aumento de la polución en el precio de las viviendas, compare con (a) y explique la diferencia.

- (c) Obtenga la estimación del modelo en (b) utilizando el teorema de la regresión particionada. Primero, guarde los residuos de una regresión de $\log(\text{nox})$ con rooms , $\log(\text{dist})$ y stratio . Segundo, haga una regresión de $\log(\text{price})$ contra los residuos. Verifique que la estimación coincide con (b).

Nota: No se preocupe por los s.e. solamente por las estimaciones puntuales.