

Trabajo Práctico 2

Modelo Lineal General y algunas extensiones

Contenidos: Modelo lineal con K variables, interpretación de los parámetros poblacionales, notación matricial, variables explicativas binarias, modelos no lineales en las variables, propiedades algebraicas de los estimadores *MCO* en notación matricial.

Fecha de entrega: jueves 8 de octubre. Ver reglas de formato y presentación en <https://econometria1unlp.com/trabajos-practicos/>.

Parte I

1. Ecuaciones de ingresos y retornos a la educación

Una rama de la economía laboral se concentra en el estudio de los determinantes de los ingresos laborales. ¿Qué hace que un individuo gane más que otro? Citando a Adam Smith en *La Riqueza de las Naciones* (libro I, Cap. 10): “Los salarios crecen con la dificultad y el costo de aprender una tarea”. Esto plantea un nexo teórico entre la educación y los salarios: la educación aumenta la productividad del trabajador y consecuentemente su salario.

Una herramienta para estudiar empíricamente la relación entre la educación y los salarios son las llamadas *ecuaciones de ingresos* o *ecuaciones de Mincer*.¹ La ecuación (1) es la especificación más básica de una ecuación de ingresos:

$$(1) \quad \ln Y_i = \beta_1 + \beta_2 \text{edu}_i + \beta_3 \text{exp}_i + \gamma \text{hombre}_i + \mu_i$$

donde Y_i son los ingresos laborales por hora o salarios horarios del trabajador i , edu_i son sus años de educación formal, exp_i sus años de experiencia laboral, hombre_i es una variable dummy que indica con 1 si se trata de un hombre y 0 en caso contrario y μ_i es un término aleatorio que supondremos que cumple con los supuestos clásicos.

En base a esto se pide:

- Obtener el efecto marginal parcial de la variable edu sobre el salario horario e interpretar su significado. Según la teoría del capital humano: ¿qué signo esperaríamos para este efecto marginal?
- Obtener analíticamente la expresión para el efecto marginal de la experiencia laboral sobre el salario horario. Interpretarlo económicamente.

¹ Mincer, J. (1974). Schooling, experience, and earnings. National Bureau of Economic Research. Columbia University Press. New York.

c) En este inciso nos vamos a concentrar en las brechas salariales entre hombres y mujeres que comparten las mismas características observables. En base a la ecuación (1), obtener la expresión que indica la diferencia promedio en el salario horario de un hombre en comparación a una mujer de iguales características. Proponer una hipótesis nula y alternativa para evaluar la ausencia de brecha salarial entre hombres y mujeres. Luego, suponer que $\gamma > 0$. ¿Es una prueba suficiente para decir que hay discriminación contra las mujeres en el mercado laboral?

2. Demanda de ocio

La siguiente ecuación es una forma simplificada del modelo utilizado por Biddle y Hamermesh (1990) para estudiar el *trade-off* entre el tiempo destinado a dormir (ocio) y el tiempo destinado al trabajo:

$$(2) \quad \text{dormir}_i = \beta_1 + \beta_2 \text{hombre}_i + \beta_3 \text{trabajo}_i + \beta_4 \text{hombre} \cdot \text{trabajo}_i + \\ + \beta_5 \text{educ}_i + \beta_6 \text{edad}_i + \beta_7 \text{edad}_i^2 + \mu_i$$

Las variables dormir_i y trabajo_i están medidas en minutos semanales, mientras que educ_i y edad_i están expresadas en años.² Utilizando la ecuación (2) se pide:

- ¿Es este un modelo lineal en parámetros? ¿Es lineal en variables? Justificar adecuadamente la respuesta.
- Obtener la expresión para el efecto marginal de la edad sobre el tiempo de sueño e interpretarlo intuitivamente. Proponer una hipótesis nula y una alternativa para evaluar si el efecto marginal de la edad sobre el sueño es constante. Suponer que $\beta_6 < 0$ y $\beta_7 > 0$, interpretar.
- Obtener la expresión que describe el trade-off entre sueño y trabajo. Proponer una hipótesis nula que postule que por cada minuto más de trabajo, los hombres duermen un minuto menos. ¿Cómo sería esa hipótesis para el caso de las mujeres?

3. Gasto en Investigación y Desarrollo (I+D)

El objetivo del siguiente modelo es conocer si la intensidad del gasto en investigación y desarrollo (I+D) aumenta con el tamaño de la firma:

$$(3) \quad ID_i = V_i^{\beta_2} \cdot \exp(\beta_1 + \beta_3 G_i + \mu_i)$$

donde la variable ID_i es el gasto en I+D como porcentaje de las ventas de la firma i , V_i mide las ventas de la firma i en millones de pesos y la variable G_i representa las ganancias como porcentaje de las ventas de dicha firma. Notar que la forma funcional no es lineal en los parámetros β_1 , β_2 y β_3 .

En base esto se pide:

² Biddle, J. E., and D. S. Hamermesh (1990), "Sleep and the Allocation of Time," *Journal of Political Economy* 98, 922–943.

- a) ¿Podríamos estimar el modelo (3) por MCO? Transformar el modelo tomando logaritmos a ambos lados de la igualdad. ¿Qué parámetros pueden estimarse a partir del modelo transformado?
- b) Utilizar el modelo transformado para obtener analíticamente el efecto parcial de un aumento en las ganancias como porcentaje de las ventas (variable G) sobre el gasto en I+D como porcentaje de las ventas (variable ID). Interpretar el significado.
- c) Suponer que estimamos el modelo transformado por MCO y obtuvimos $\hat{\beta}_2 = 1.3$ ¿Cuál es la interpretación económica de ese valor?

Parte II ³

4. Notación matricial: modelo con 3 variables

Considerar nuevamente el modelo de ahorro S e ingresos I de las familias propuesto en el ejercicio 3 del Trabajo Práctico 1. Suponer que otro de los determinantes del nivel de ahorro en el hogar puede ser el número de hijos a cargo. El ahorro queda definido por la siguiente ecuación:

$$S_i = \alpha + \beta I_i + \delta N_i + \mu_i \quad i = 1, \dots, n$$

en donde S_i e I_i son el ahorro e ingreso de las familias respectivamente (medidos en pesos) y N_i es el número de niños en el hogar. Los datos disponibles son los siguientes:

| Hogar (i) | Ahorro (S) | Ingreso (I) | Niños (N) |
|------------------|-------------------|--------------------|------------------|
| 1 | 97 | 141 | 1 |
| 2 | 46 | 145 | 0 |
| 3 | 125 | 209 | 1 |
| 4 | 81 | 132 | 2 |
| 5 | 68 | 137 | 3 |
| 6 | 147 | 205 | 1 |
| 7 | 12 | 61 | 4 |
| 8 | 84 | 136 | 1 |
| 9 | 105 | 175 | 0 |
| 10 | 93 | 140 | 2 |
| 11 | 15 | 41 | 1 |
| 12 | 112 | 158 | 0 |
| 13 | 14 | 42 | 2 |
| 14 | 61 | 78 | 2 |
| 15 | 59 | 130 | 1 |

Se pide:

³ Puede utilizar calculadora o planilla de cálculo para hacer sumas o productos de datos, pero las operaciones con matrices deben hacerse *a mano*.

- a) Definir las matrices que intervienen en el cómputo de los estimadores de mínimos cuadrados ordinarios de este modelo. Especificar cuáles son las matrices, cuáles son sus dimensiones y qué elementos contiene cada una de ellas.
- b) Utilizando las matrices, computar los estimadores de mínimos cuadrados ordinarios del modelo.
- c) Interpretar el significado económico de los coeficientes estimados.

5. Caso particular: $n = k$

En el modelo lineal con k variables $Y = X\beta + \mu$, en donde Y es un vector de n observaciones, X una matriz de K variables explicativas y n observaciones y μ un vector que contiene los n términos aleatorios.

Mostrar que si el número de observaciones es igual al número de variables explicativas (incluyendo la constante), entonces la suma de los residuos al cuadrado es igual a cero y el R^2 igual a 1. Explicar.

[Ayuda: notar que cuando $n = K$, X es una matriz cuadrada]

6. Estimación computacional: modelo no lineal en las variables

La Economía Ambiental es una rama de la economía que se ocupa del estudio de las decisiones económicas de los agentes y sus efectos sobre el medio ambiente. El índice de contaminación más utilizado es el nivel de emisiones de dióxido de carbono.

Un resultado interesante de esta literatura (conocido como Curva de Kuznets Ambiental) es que existe una relación con forma de U invertida entre el nivel de contaminación y el desarrollo económico de un país; es decir, a medida que la economía se desarrolla el nivel de polución ambiental se incrementa hasta un cierto punto en donde el proceso se revierte y el crecimiento económico comienza a estar asociado con menores niveles de contaminación.

Por otro lado, varias teorías destacan la relevancia de la educación sobre el grado de conciencia social en las cuestiones ambientales, en el sentido de que una población más educada tomaría decisiones que generan menores niveles de contaminación del medio ambiente.

La siguiente ecuación trata de capturar las relaciones mencionadas anteriormente:

$$(6) \quad \ln CONT_i = \beta_1 + \beta_2 PBI_i + \beta_3 PBI_i^2 + \beta_4 EDU_i + \mu_i$$

donde $CONT_i$ es la emisión de dióxido de carbono del país i (en toneladas métricas per cápita), PBI_i es el producto bruto interno per cápita medido en dólares del país i , EDU_i es el promedio de años de educación del país i y μ_i es un término aleatorio que supondremos que cumple con los supuestos clásicos.

El archivo *polucion2006.dta* contiene datos para 150 países en el año 2006 sobre las siguientes variables:

- **diox**: nivel de contaminación, medido como la intensidad en las emisiones de dióxido de carbono (CO2) en Kg, por cada Kg de petróleo utilizado.

- *pbi*: PBI per cápita en miles de dólares corrientes.
- *edu*: años promedio de educación.

Utilizando esta información se pide completar los siguientes incisos:

- a) Usando Stata estimar por MCO los parámetros del modelo planteado por la ecuación (6)
- b) Describir el efecto marginal estimado de la educación promedio sobre el nivel de contaminación. ¿Son los años de educación promedio relevantes para explicar la emisión de dióxido de carbono de un país? Realizar la prueba contra una hipótesis alternativa bilateral. Especificar el nivel de confianza con el que se realiza el test.
- c) Evaluar la hipótesis nula de que la relación entre el nivel de contaminación y el PBI per cápita de los países es lineal contra la alternativa de que la relación entre ambas variables es cóncava ($\beta_3 < 0$). Utilizar un nivel de confianza del 95%.
- d) Utilizando los coeficientes estimados, construir un gráfico del valor de contaminación predicho (en logaritmos) y el PBI per cápita para un país con 12 años de nivel de educación promedio. Describir lo observado.

7. Estimación de la relación entre el ingreso laboral y sus determinantes.

En este ejercicio se estimará la relación entre el ingreso laboral mensual y determinadas características observables de los trabajadores. La base *EPHC_2019_s2.dta* contiene microdatos de la Encuesta Permanente de Hogares de Argentina del año 2019, semestre 2. La información corresponde a 8,374 personas empleadas del Gran Buenos Aires que reportaron información sobre el número de horas que trabajaron en la semana anterior a la encuesta, el valor de su ingreso laboral mensual expresado en pesos, y sus años de educación.

Se pide:

- a) Escribir formalmente la ecuación de un modelo lineal en parámetros donde la variable a explicar es el logaritmo del ingreso laboral mensual y las variables explicativas son las horas de trabajo semanales, el cuadrado de las horas de trabajo semanales y los años de educación.
- b) Estimar por MCO el modelo anterior. Interpretar económica y estadísticamente el efecto parcial de la variable años de educación. Especificar el nivel de confianza con el que se realiza el test.
- c) Obtener el efecto marginal de la variable horas de trabajo y evaluar la hipótesis de que la relación entre el ingreso laboral mensual y las horas de trabajo es lineal contra una hipótesis alternativa bilateral. Utilizar un nivel de confianza del 99%.