

## Material para las clases teóricas virtuales

### Clase 4: El modelo lineal con K variables. Interpretación y usos. ([Link](#) a lista de reproducción)

1. Presentación en pdf para imprimir disponible en [este link](#)
2. Videos (son 4 videos en total):
  - [Video 1 – El modelo lineal general](#)  
Se presenta y discute la interpretación del modelo lineal general o modelo con K variables.

#### Preguntas guía para el video 1:

- a) En el modelo lineal general, dos variables explicativas pueden estar fuertemente correlacionadas y eso no viola el supuesto de no multicolinealidad perfecta. ¿Por qué?
- b) ¿Por qué en el modelo lineal general ninguna de las variables explicativas puede ser constante?
- c) Al final del video la profesora dice 2 veces algo incorrecto. Encontrar el error (ver a partir del minuto 13:00).

- [Video 2 – Variables binarias](#)

Se explica el uso y la interpretación de las variables binarias, tanto para representar fenómenos binarios como fenómenos con categorías múltiples.

#### Preguntas guía para el video 2:

- a) La regla práctica nos dice que si queremos incorporar información de un fenómeno binario debemos incluir solo una dummy al modelo. ¿Qué sentido tiene esta regla?
- b) En el modelo que incorpora las regiones (filmina 17 y siguientes), escribir el valor esperado de W para cada grupo posible (mujeres en cada una de las regiones y hombres en cada una de las regiones). Luego responder: ¿cuál es la diferencia en el valor esperado del salario entre trabajadores de la región 1 y de la región 3, si todas las demás características se mantienen constantes? ¿y entre la región 2 y la región 3? ¿y entre la región 1 y la región 2? ¿cuál es la diferencia en el valor esperado del salario entre un trabajador hombre de la región 2 y una trabajadora mujer de la región 1? ¿sabemos el signo de esa diferencia? En cada caso pensar bien cuáles son las demás características que se mantienen constantes.
- c) Ver la página 24 de las filminas, que no aparece en el video.

- [Video 3 – No linealidades en variables](#)

Se presentan casos donde las variables aparecen de manera no lineal en el modelo y se discute su interpretación: variables en logaritmos, variables cuadráticas e interacciones entre variables.

#### Preguntas guía para el video 3:

- a) ¿Por qué en la filmina 30 dice que es “por construcción” que el modelo lineal en X predice un efecto marginal constante de X?
- b) Considerar los siguientes dos modelos

$$(1) Y_i = \beta_1 + \beta_2 X_i + \delta \text{hombre}_i + \mu_i$$

$$(2) Y_i = \beta_1 + \beta_2 X_i + \delta \text{hombre}_i + \gamma (X_i \times \text{hombre}_i) + u_i$$

Notar que (2) es el modelo de la filmina 37 y (1) es una versión simplificada, que no incluye el término de interacción.

- Obtener la expresión del efecto marginal de  $X$  sobre el valor esperado de  $Y$  en cada modelo.
  - Alguien estimó el modelo (1) por MCO y sacó la siguiente conclusión: “Estos resultados dicen que el efecto marginal de  $X$  sobre  $Y$  no depende del género”. ¿Qué opinás?
- c) ¿Qué signo deberían tener los parámetros para que el gráfico de la filmina 39 sea válido?

○ **Video 4 – Ejemplo empírico**

Se discute un ejemplo sobre la estimación de un modelo de los determinantes de los salarios usando datos de la Encuesta Permanente de Hogares de Argentina. En este ejemplo se incorporan todos los elementos aprendidos en la clase 4: variable dependiente en logaritmos, variables explicativas dummy, cuadráticas e interacciones.

Preguntas guía para el video 4:

- a) En el video se habla del p-valor (a partir del minuto 3:14). La salida de regresión de STATA reporta el p-valor para evaluar la significatividad estadística de la variable explicativa cuando la hipótesis alternativa es a dos colas, es decir,  $H_0: \beta_2 = 0$  versus  $H_A: \beta_2 \neq 0$ . Con relación a esto:
- Primero, asegurarse de entender qué es, cómo se computa y cómo puede usarse el p-valor al realizar un test de hipótesis (reparar Estadística II). Puede ser útil ayudarse con un gráfico de la distribución  $T$  bajo la hipótesis nula, donde se marquen las zonas de rechazo y de aceptación de la  $H_0$ , el  $t$  observado y el p-valor.
  - Después pensar cómo se puede usar el p-valor que aparece en la salida de regresión de STATA si la hipótesis alternativa es a una sola cola, por ejemplo, si quisiéramos evaluar  $H_0: \beta_2 = 0$  versus  $H_A: \beta_2 > 0$ .
- b) Hacer un gráfico que muestre la regresión estimada en función de la edad, manteniendo fijas las demás variables. Poner en el eje vertical el logaritmo del salario y en el eje horizontal la edad. Explicar cómo va cambiando el efecto marginal de la edad y para qué edad se estima que cambia de signo.
- c) En la filmina 44 dice “el salario de un hombre es, en promedio, aproximadamente un 7% ( $0.0712 \times 100\%$ ) mayor que el de una mujer de la misma edad y con la misma educación”. Suponer ahora que el coeficiente estimado fuera igual a 0.424 en lugar de 0.0712. ¿Cuál sería la brecha salarial estimada entre hombres y mujeres, dado todo lo demás?
- d) ¿Cómo puede evaluarse la hipótesis de que el retorno a la educación no difiere entre hombres y mujeres en el modelo de la filmina 47?