Econometría I

2. Modelo de Regresión Lineal Simple (MRLS)

José Carlos Tello

Universidad Americana

Clase 2 — 11 de agosto
Segundo Semestre 2023

Modelo de Regresión Lineal Simple (MRLS)

Introducción

- Dado que en muchas situaciones no es posible realizar experimentos con asignación aleatoria (lo ideal) entonces buscamos nuevas rutas para identificar relaciones causales.
- Una de ellas es el Modelo de Regresión Lineal Múltiple (MRLM) pero antes de eso es útil comprender algunos conceptos básicos.
- A pesar de sus serias limitaciones, el Modelo de Regresión Lineal Simple (MRLS) permite entender el método de estimación y comprender el proceso de inferencia estadística.
- Sin embargo, debemos tener siempre presente que el MRLS es equivalente al coeficiente de correlación simple, cuya capacidad para identificar relaciones causales en entornos económicos hemos cuestionado en la clase 1.

Modelo de Regresión Lineal Simple (MRLS)

Introducción

- Nos interesa explicar el comportamiento de cierta variable "y" a partir de cambios o movimientos de otra "x".
- Ejemplos:

y	**x**	Investigación
Salarios	Años de educación	Card (2001)
Crecimiento económico	Calidad de instituciones	Acemoglu, Johnson y Robinson (2001)
Empleo	Salario mínimo	Card y Krueger (1994)
Prima de riesgo de un activo	Prima de riesgo del mercado	Kothari, Shanken y Sloan (1995)
Desempleo	Inflación	Galí, Gertler y López-Salido (2005)

Modelo de Regresión Lineal Simple (MRLS)

Introducción

- Con frecuencia nos interesa conocer el valor promedio de "y" para cada potencial valor de "x", es decir, $E[y|x]$ (media condicional de y en x)
- Si la variable aleatoria y es continua entonces $E[y|x] \equiv \int y \ f(y|x)dx$.
- Si es discreta entonces $E[y|x] \equiv \sum yf(y|x)$.
- Sin embargo, **nunca vamos a conocer** $f(y|x)$.
- ¿Qué se puede hacer? ¡Una aproximación!
- Para hacer esto necesitamos empezar con algunos supuestos.

Modelo de Regresión Lineal Simple (MRLS)

Definición

- Supongamos que la relación entre "y" y "x" es **lineal**.

$$y = \beta_0 + \beta_1 x + u. \tag{1}$$

- Incluimos el término "u", ahora denominado error o perturbación, para mostrar que la relación lineal entre "y" y "x" no es exacta.
- Componente "u" contiene factores, distintos a "x", que afectan o determinan los valores de "y".
- La variable de nuestro interés "y" se le conoce como variable dependiente, explicada o de respuesta.
- Variable "x" se le conoce como variable independiente[1] o explicativa.

---

[1] Cuidado, no debe entenderse que las variables aleatorias x e y son independientes entre sí.

Modelo de Regresión Lineal Simple (MRLS)

Características

- La linealidad del modelo es con respecto a los parámetros.
- Así el efecto de "x" sobre "y" es constante: $\frac{\partial y}{\partial x} = \beta_1$.
- **Advertencia**. Si bien, toda derivada parcial toma el resto de factores constantes (*ceteris paribus*), aún no tenemos las condiciones necesarias para discutir el concepto de causalidad.
- Las siguientes modificaciones a la ecuación (1) son también lineales con respecto a los parámetros:

$$
\begin{gathered}
y = \alpha_0 + \alpha_1 \frac{1}{x} + u \ ; \\
\ln y = \gamma_0 + \gamma_2 \ln x + u \ ; \\
y = \theta_0 + \theta_1 \sqrt{x} + u.
\end{gathered}
$$

- Tarea: Verificar que los impactos de la variable explicativa sobre la variable de respuesta es constante.

Modelo de Regresión Lineal Simple (MRLS)

Características

- Si tomamos la función determinística $f(x) = \beta_0 + \beta_1 x$ y la graficamos en el plano (X,Y) decimos que β_1 es la pendiente de la función.
- Mientras que β_0 es el intercepto.
- Una crítica al supuesto de un modelo lineal es que no siempre las variables van a mostrar una relación funcional tan sencilla.
- Sin embargo, siempre es posible "linealizar" toda relación entorno a un valor de x e y.[2]

[2] En este proceso se puede utilizar series de Taylor de orden n, si n=1 esta linealización coincide con la aplicación de la función logaritmo natural.

Modelo de Regresión Lineal Simple (MRLS)

Media Condicional y Función de Regresión Poblacional

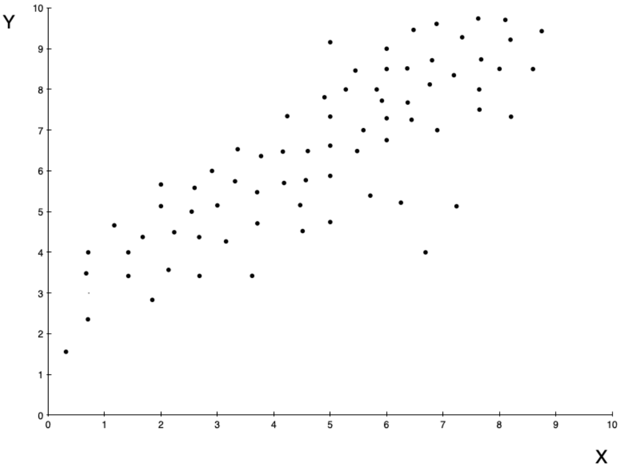
- Hacer una regresión tiene como objetivo "decir cosas" de "y" a partir de otra variable "x".
- Si la verdadera relación (poblacional) es el modelo (1) y le tomamos la **media condicional** en "x" obtenemos:

$$E\left[y|x\right] = \beta_0 + \beta_1 x + E\left[u|x\right]. \tag{2}$$

- Un supuesto razonable es considerar $E\left[u|x\right] = 0$.
- Es decir, dado un valor de "x", el valor promedio de los errores es cero o que "u" es media independiente de "x".
- La ecuación $E\left[y|x\right] = \beta_0 + \beta_1 x$ se le conoce como **Función de Regresión Poblacional (FRP)**.

Modelo de Regresión Lineal Simple (MRLS)

Media Condicional y Función de Regresión Poblacional



- La FRP es una aproximación de la media condicional $E[y|x]$.
- Esta aproximación está condicionada a los supuestos:
  - S1. Linealidad en los parámetros.
  - S2. $E[u|x] = 0$.
- El supuesto $E\left[u|x\right] = 0$ dice que la **media de $\boldsymbol{u}$ no depende de $\boldsymbol{x}$** (tanto en forma lineal como no-lineal).
- Del **S2** se deduce dos cosas importantes:
  - $E\left[u\right] = 0$. El valor promedio poblacional de los errores es cero, para cualquiere valor de x.
  - $E\left[xu\right] = 0$. La covarianza poblacional entre la explicativa y el error es cero, es decir, no deberían tener una relación lineal.

Modelo de Regresión Lineal Simple (MRLS)

Media Condicional y Función de Regresión Poblacional

- **Advertencia**. Si $E[u] = 0$ y $E[xu] = 0$ no se puede deducir que $E[u|x] = 0$.
- La relevancia del supuesto $(E[u|x] = 0)$ es que permite dividir la media condicional en una parte sistémica $(\beta_0 + \beta_1 x)$ y una parte no sistémica $(E[u|x])$.
- La media condicional y la FRP son conceptos teóricos.
- **Nunca** conoceremos los valores de los parámetros β_0 y β_1.
- Pero si disponemos de una muestra aleatoria con n observaciones $\{(x_i, y_i) : i = 1, ..., n\}$ entonces podemos **estimar** los parámetros de la media condicional.

Referencias

1. Acemoglu, D., Johnson, S., y Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. American Economic Review, 91(5), 1369-1401.
2. Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica, 69(5), 1127-1160.
3. Card, D., y Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review, 84(4), 772-93.
4. Galí, J., Gertler, M., y López-Salido, J. D. (2005). Robustness of the estimates of the hybrid New Keynesian Phillips curve. Journal of Monetary Economics, 52(6), 1107-1118.
5. Kothari, S. P., Shanken, J., y Sloan, R. G. (1995). Another look at the cross-section of expected stock returns. The Journal of Finance, 50(1), 185-224.