

# Regresión Lineal y sus Aplicaciones

## Escuela de Invierno en Métodos y Análisis de Datos UCU-DCSP

Santiago López Cariboni  
*Departamento de Ciencias Sociales y Políticas*  
*Universidad Católica del Uruguay*  
santiago.lopez@ucu.edu.uy

### Descripción

Este curso se concentra la comprensión del modelo de regresión lineal y su utilidad para la realización de inferencias causales, así como sus limitaciones. El curso comienza revisando los problemas que enfrenta el investigador para realizar inferencias sobre la base de datos observacionales. Luego introduce la teoría estadística básica y la estimación por medio de regresión lineal como una herramienta adecuada para la inferencia bajo determinadas circunstancias. Los participantes van a adquirir el conocimiento de los supuestos que sustentan la estimación mediante mínimos cuadrados, las consecuencias de sus violaciones y las soluciones existentes a esos problemas. Asimismo, los participantes ganarán un fuerte entrenamiento práctico en programar y estimar sus resultados utilizando el paquete de código abierto R (<http://www.r-project.org/>). Participantes interesados en trabajar en STATA también son bienvenidos. Las clases de laboratorio constan de una serie de ejercicios prácticos replicando resultados publicados en revistas científicas y analizando alternativas de especificación de esos mismos modelos.

### Objetivos

Luego del curso los estudiantes habrán cumplido con los siguientes objetivos:

- Comprender los fundamentos de la regresión lineal, sus supuestos, y los contextos en los cuales es útil para la realización de inferencias.
- Utilizar R (o STATA) para generar y presentar sus resultados de regresión.
- Interpretar resultados de regresiones, tales como coeficientes, insignificancia estadística, incertidumbre y calidad del modelo.
- Analizar y corregir por potenciales violaciones a los supuestos de la regresión lineal.

### Prerrequisitos

Los participantes deben tener conocimientos básicos de estadística descriptiva y sentirse cómodos con álgebra lineal. Aunque resulte de utilidad, previo conocimiento en R o STATA no es estrictamente necesario.

## Lecturas

### Teóricas

Joshua D. Angrist and Jörn-Steffen Pischke. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

Andrew Gelman and Jennifer Hill. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Charles F. Manski. (2007). *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA.

Jeffrey Wooldridge M. (2009). *Introductory Econometrics. A Modern Approach*. South-Western.

### Libros para R y STATA

Cristopher F. Baum. (2006). *An introduction to modern econometrics using Stata*. Stata Press.

John Fox and Sanford Weisberg (2011). *An R companion to applied regression*. Sage.

Christian Kleiber and Achim Zeileis (2008). *Applied econometrics with R*. Springer.

## Esquema del curso

**Día 1: Principios de la inferencia estadística.** Inferencia descriptiva, inferencia predictiva, e inferencia causal. Variables, distribuciones, y expectativas. Identificación, estimación, intervalos de confianza, y test de hipótesis.

### Lecturas:

Wooldridge (2009), Apéndices (B - C).

Gelman y Hill (2007). Cap. 2.

Angrist y Pischke (2009). Cap. 2.

Manski (2007) "Introduction" y Cap. 7.

**Laboratorio:** Tests simples ( $\chi^2$ , F, t), comparación de medias y desvíos estándar.

**Día 2: Regresión lineal.** Regresión bivariada: principio de los mínimos cuadrados, parámetros, y componentes del análisis de regresión. Regresión multivariada: estimación.

### Lecturas:

Wooldridge (2009), Cap. 2-3.

Gelman y Hill. Cap. 3-4.

Kleiber y Zeileis (2008). Cap. 3.

Baum. (2006). Cap. 4.

**Laboratorio:** Calculo manual de coeficientes e incertidumbre. Aplicaciones en contextos experimentales. Estimación en R y STATA.

**Día 3: Regresión lineal multivariada: inferencia y propiedades asintóticas de OLS.** Supuestos Gauss-Markov. Propiedades de mínimos cuadrados en muestras finitas y no finitas. Interpretación de resultados de regresión.

**Lecturas:**

Wooldridge (2009). Caps. 4-5.

Angrist y Pischke (2009). Cap. 3.

Kleiber y Zeileis (2008). Cap. 3.

Baum. (2006). Cap. 4.

**Laboratorio:** Ejercicios de replicación de resultados publicados y re-análisis de los datos.

**Día 4: Problemas de especificación.** Linealidad, información cualitativa (variables “dummy”), efectos de interacción.

**Lecturas:**

Wooldridge (2009). Caps. 6-7.

Gelman y Hill. Cap. 3-4.

Brambor, Thomas, Clark, William Roberts and Matt Golder (2006). “Understanding Interaction Models: Improving Empirical Analysis”, *Political Analysis* 14, 63-82.

**Laboratorio:** Ejercicios de especificación: transformación de los datos, no linealidades, programación de interacciones y visualización de efectos marginales.

**Día 5: Violaciones de supuestos Gauss-Markov y soluciones.** Heteroscedasticidad, sesgo por variables omitidas, multicolinealidad, outliers, correlación transversal. Introducción a endogeneidad. Compresión de Tests disponibles, y soluciones.

**Lecturas:**

Wooldridge (2009). Caps. 8-9 y 15.

Angrist y Pischke (2009). Cap. 4.

**Laboratorio:** Bateria de tests en R y STATA. Estimación con variables instrumentales.