

ITAM
Departamento de Ciencia Política

Metodología Política Avanzada

Otoño 2018
Viernes 12:00-15:00
Río Hondo, Salón CC201

Adrián Lucardi
adrian.lucardi@itam.mx

Oficina: Martes 15:00-16:30 y Jueves 16:00-17:30 o previa cita

En este curso veremos los rudimentos del análisis de información cuantitativa en ciencias sociales, con especial énfasis en (a) la *manipulación* y *visualización* de datos; (b) la *especificación* de modelos estadísticos paramétricos; y (c) la *interpretación* de los resultados de dichos modelos. El contenido del curso se centrará en cuatro grandes temas que constituyen la base de la investigación aplicada en ciencias sociales en general y en ciencia política en particular:

- (i) Especificación e inferencia con modelos de regresión;
- (ii) Estimación e interpretación de interacciones entre variables;
- (iii) Estimación e interpretación de modelos con variables dependientes discretas; y
- (iv) Manipulación y análisis de datos agrupados, incluyendo datos de panel y modelos con efectos fijos y aleatorios.

En términos pedagógicos, partiremos de la idea de que la estadística no es una técnica ni una metodología sino un lenguaje, y la mejor forma de aprender un lenguaje es practicándolo. Por lo tanto, si bien habrá comparativamente pocas lecturas, deberás completar numerosos ejercicios prácticos – tanto individuales como colectivos – *fuera del horario de clases*. Además, tanto el material teórico como los ejercicios de manipulación y análisis de datos son acumulativos, por lo que es indispensable mantenerse al día.

No necesitas estar familiarizado con ningún lenguaje de programación. Sin embargo, en la presentación del material teórico daré por sentado que conoces los rudimentos de teoría de probabilidades y modelos de regresión por mínimos cuadrados ordinarios; de no ser así, debes ponerte al día por tu cuenta. Asimismo, si bien a lo largo del curso discutiremos cuándo es posible interpretar los resultados estadísticos en términos causales, el foco del curso estará puesto en la especificación e inferencia de modelos estadísticos, no en el problema de la causalidad como tal.

Calificación

- **Participación en clase (10%).** Debes leer la bibliografía obligatoria y participar activamente en las discusiones. Los días que haya trabajo en grupo no está permitido faltar a riesgo de recibir una calificación de 0.

- **Tareas (40%).** A lo largo de la clase tendrás que entregar ocho tareas. Puedes trabajar en grupo para responderlas, pero cada tarea debe ser completada – y será evaluada – de forma estrictamente individual. Debes entregar las tareas *antes* del comienzo de la siguiente clase. Cada tarea será calificada como (i) entregada (5% de la calificación final); (ii) entregada incompleta (3%); o (iii) no entregada (0%).
- **Replicación en grupo (30%).** Habrá tres sesiones de trabajo grupal orientadas a reproducir los resultados de un artículo recientemente publicado en una revista académica de ciencia política. Los detalles los discutiremos en clase. Cada grupo entregará los resultados de la replicación *antes* del comienzo de la siguiente clase. Cada trabajo contará el 10% de la calificación final.
- **Trabajo final (20%).** Durante la semana de finales deberás entregar un trabajo individual consistente en (a) la replicación de los resultados de un artículo recientemente publicado en una revista académica de ciencias sociales; ó (b) el análisis de datos que quieras emplear en tu tesis. Los detalles los discutiremos en clase, pero en ambos casos el proyecto debe ser previamente aprobado por el profesor.

Lecturas

Buena parte de las discusiones teóricas estarán basadas en el siguiente libro:

- Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 5th edition, 2012.

Sin embargo, algunas discusiones teóricas estarán basadas en otros textos, listados más abajo. Si lo deseas, puedes complementar las lecturas obligatorias con alguna de las siguientes lecturas, ordenadas de menor a mayor en términos de dificultad:

- Joshua D. Angrist and Jörn-Steffen Pischke. *Mastering 'Metrics. The Path from Cause to Effect*. Princeton University Press, Princeton, 2015.
- David M. Diez, Christopher D. Barr, and Mine Çetinkaya-Rundel. *OpenIntro Statistics: Third Edition*. CreateSpace Independent Publishing Platform, Lexington, KY, 3rd edition, 2015.
- Peter Kennedy. *A Guide to Econometrics*. Wiley-Blackwell, Malden, MA, 6th edition, 2008.
- John Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage, Los Angeles, 3rd edition, 2015.
- Joshua Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton University Press, Princeton, 2009.

Todas las lecturas se encuentran disponibles en [Google Drive](#).

Software

Cada vez que entregues un trabajo grupal o individual, deberás adjuntar el código correspondiente en R ó Stata. No se aceptarán otros lenguajes de programación. No obstante, ten en cuenta que la

mayoría de los ejemplos de código que distribuiré a lo largo del curso serán en R; por lo que a menos que ya estés familiarizado con Stata, te recomiendo hacer tus tareas en R.

Puedes descargar R de manera gratuita del siguiente link: <https://cran.itam.mx/>. En ese caso, te recomiendo que lo uses en conjunto con RStudio (<https://www.rstudio.com/>). Si quieres usar Stata, tendrás que averiguar la forma de instalarlo en tu computadora, o bien trabajar en las salas de cómputo del ITAM.

Además de los ejemplos de código distribuidos en clase, la manera más efectiva de aprender a programar es buscar código de otros autores en internet y adaptarlo a tus propias necesidades. También puedes trabajar por tu cuenta con alguno de los siguientes libros, disponibles online y/o en la carpeta de [Google Drive](#) del curso:

- Para R: John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks (CA), 2nd edition, 2011.
Paul Teetor. *R Cookbook*. O'Reilly, Sebastopol (CA), 2011.
Para visualización de datos, puedes consultar el manuscrito del siguiente libro: Kieran Healy. *Data Visualization: A Practical Introduction*. Princeton University Press, Princeton and Oxford, forthcoming. Disponible en <http://socviz.co/>.
- Para Stata: A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics Using Stata [Revised Edition]*. Stata Press, College Station (TX), [2009] 2010.

Integridad académica

Según el artículo 5bis del Reglamento de Titulación del ITAM, “la reproducción parcial o total de textos o productos ajenos, de autor conocido o anónimo, publicados por cualquier medio o inéditos, sin entrecomillar los pasajes o elementos reproducidos o sin hacer la referencia bibliográfica pertinente,” constituye plagio o fraude académico. Las citas a otros autores son bienvenidas, pero deben ser reconocidas como tales. El uso de fragmentos y/o traducciones literales de textos ajenos sin la debida atribución será causal para reprobar la materia y merecerá la amonestación oficial al alumno por parte de la Dirección Escolar del Instituto.

Calendario

Sesión 1 (Agosto/17). Introducción

- Presentación del curso.
- *Teoría*: Wooldridge, sec. 1.3.
 - tipos de datos
- *Práctica*: introducción a R (o Stata):
 - instalación y configuración
 - importando datos
 - estadística descriptiva

- histogramas / gráficas de densidad
- *scatterplots*
- Consigna para la *Tarea #1*.

Sesión 2 (Agosto/24). Manipulación de datos

- Entrega y revisión de la *Tarea #1*.
- *Teoría*: tipos de variables.
- *Práctica*: manipulación básica de variables en Stata o R:
 - creando nuevas variables
 - creación y manipulación de distintos tipos de variables: continuas, *dummies*, factores y factores ordenados
 - *scatterplots* de variables continuas y factores: *boxplots*
 - manipulando *missing values*
 - seleccionando subconjuntos de datos
 - calculando estadísticas descriptivas por factores
- Consigna para la *Tarea #2*.

Agosto/31. APSA. No hay clase. Pero:

- *Lectura*:
 - (1) Jeff Gill. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly*, 52(3):647–674, 1999.
 - (2) Andrew Gelman and David Weakliem. Of Beauty, Sex, and Power: Statistical Challenges in Estimating Small Effects. *American Scientist*, 97:310–316, 2009.
- *Práctica*: continuación de la *Tarea #2*:
 - visualización de datos: series de tiempo, partidos y votos
 - creando y manipulando subconjuntos de datos

Sesión 3 (Septiembre/7). Regresión simple

- Entrega y revisión de la *Tarea #2*.
- *Teoría*: Wooldridge, secs. 2.1-2.4.
 - intuición
 - efectos marginales y predicciones
 - transformando variables
 - regresión contra una constante
- *Práctica*: visualización
 - *scatterplots* con línea de regresión
 - coeficientes e intervalos de confianza
- Consigna para la *Tarea #3*.

Sesión 4 (Septiembre/14). Regresión múltiple

- Entrega y revisión de la *Tarea #3*.
- *Teoría*: Wooldridge, secs. 3.1-3.2 y 7.2-7.3
 - interpretación
 - variables independientes continuas y discretas
 - multicolinealidad y multicolinealidad perfecta
 - variables irrelevantes
 - predicciones
- *Lecturas optativas*:
 - (1) Christopher H. Achen. Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong. *Conflict Management and Peace Science*, 22(4):327–339, 2005.
 - (2) Kevin A. Clarke. The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science*, 22(4):341–352, 2005.
- *Práctica*: visualización
 - *scatterplots* con línea de regresión
 - múltiples coeficientes e intervalos de confianza
 - transformando regresión múltiple en bivariada
- Consigna para la *Tarea #4*.

Sesión 5 (Septiembre/21). Inferencia

- Entrega y revisión de la *Tarea #4*.
- *Teoría*: Wooldridge, secs. 3.1-3.2 y 8.1-8.2.
 - heteroscedasticidad: intuición
 - errores robustos y/o agrupados
 - significancia estadística vs. significancia *sustantiva*
- *Lecturas optativas*:
 - (1) Carlisle Rainey. Arguing for a Negligible Effect. *American Journal of Political Science*, 58(4):1083–1091, 2014.
 - (2) Justin Esarey and Andrew Menger. Practical and Effective Approaches to Dealing with Clustered Data. *Political Science Research and Methods*, forthcoming.
 - (3) Charles Wheelan. *Naked Statistics: Stripping the Dread from the Data*. W. W. Norton & Company, New York, 2013, cap. 6.
 - (4) Christopher H. Achen and Larry M. Bartels. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press, Princeton, 2016, pp. 118-28.
 - (5) Anthony Fowler and Andrew B. Hall. Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence. *The Journal of Politics*, forthcoming.
 - (6) Christopher H. Achen and Larry M. Bartels. Statistics as If Politics Mattered: A Reply to Fowler and Hall. *The Journal of Politics*, forthcoming.

- (7) Anthony Fowler and Andrew B. Hall. Politics as if Evidence Mattered: A Reply to Achen and Bartels. Unpublished manuscript, 2018.

Sesión 6 (Septiembre/28). Interacciones

- *Teoría:* Thomas Brambor, William R. Clark, and Matt Golder. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, 14(1):63–82, 2006.
 - intuición y álgebra
 - efectos marginales
 - cómo predecir valores
- *Lecturas optativas:*
 - (1) William D. Berry, Jacqueline H. R. DeMeritt, and Justin Esarey. Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential? *American Journal of Political Science*, 54(1):248–266, 2010.
 - (2) William D. Berry, Matt Golder, and Daniel Milton. Improving Tests of Theories Positing Interaction. *The Journal of Politics*, 74(03):653–671, 2012.
 - (3) Jens Hainmueller, Jonathan Mummolo, and Yiqing Xu. How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. *Political Analysis*, forthcoming.
 - (4) Andrew Gelman, “[You need 16 times the sample size to estimate an interaction than to estimate a main effect](#)”.
- *Práctica:*
 - modelos con interacciones
 - cálculo de errores estándar (+ robustos y agrupados)
 - visualización de los resultados
- Consigna para la *Tarea #5*.

Sesión 7 (Octubre/5). Replicación (1)

- Entrega y revisión de la *Tarea #5*.
- *Lectura:* texto a replicar (a definir).
- *Replicación:* regresión múltiple + interacciones.

Sesión 8 (Octubre/12). Variables dependientes discretas

- Entrega y revisión de la *Replicación #1*.
- *Teoría:* Fox, sec. 14.1; Wooldridge, sec. 7.5; y Walter Sosa Escudero, “[Regresión logística ¿Para todos, algunos o nadie?](#)”
 - *logit/probit*: intuición y visualización
 - coeficientes y efectos marginales
 - predicción
 - ¿*logit/probit* o LPM?

- *Lecturas optativas:*
 - (1) Wooldridge, sec. 17.1.
 - (2) Dave Giles, “Another Gripe About the Linear Probability Model” y “Yet Another Reason for Avoiding the Linear Probability Model”.
 - (3) Steffen Pischke, “Probit Better than LPM?”
 - (4) Jed Friedman, “Whether to Probit or to Probe It: In Defense of the Linear Probability Model”.
- *Práctica:*
 - corriendo modelos con variables dependientes discretas
 - visualización de los resultados
 - cálculo de efectos marginales

Sesión 9 (Octubre/19). Efectos fijos y aleatorios

- *Teoría:* Andrew Bell and Kelvyn Jones. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, 3(1):133–153, 2015.
 - Variación “*within*” vs. “*within*”
 - efectos fijos: intuición, interpretación y supuestos
 - efectos aleatorios: intuición, interpretación y supuestos
 - predicciones
- *Lecturas optativas:*
 - (1) William H. Greene. Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models? *Political Analysis*, 19(2):135–146, 2011.
- *Práctica:*
 - manipulación y visualización de datos agrupados
 - corriendo modelos con efectos fijos: estrategias alternativas
 - ¿fijos o aleatorios? comparando resultados
- Consigna para la *Tarea #6*.

Sesión 10 (Octubre/26). Replicación (2)

- Entrega y revisión de la *Tarea #6*.
- *Lectura:* texto a replicar: Lucas Leemann and Isabela Mares. The Adoption of Proportional Representation. *The Journal of Politics*, 76(2):461–478, 2014.
- *Replicación:* variables dependientes discretas + datos agrupados.

Noviembre/2. Asueto. No hay clase

Noviembre/9. Latin American PolMeth. No hay clase. Pero:

- Entrega de la *Replicación #2*.
- Base de datos para el trabajo final.

Sesión 11 (Noviembre/16). Datos de panel

- *Teoría*: Wooldridge, cap. 13 y sec. 14.1.
 - supuestos y estructura
 - interpretación
 - variables dependientes discretas
- *Práctica*:
 - manipulación de datos de panel: *lags* y *leads*
 - corriendo modelos e interpretando resultados
 - visualización gráfica: complicada, pero aún posible
- Consigna para la *Tarea #7*.

Sesión 12 (Noviembre/23). Diferencia-en-Diferencias

- Entrega y revisión de la *Tarea #7*.
- *Teoría*: Joshua Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton University Press, Princeton, 2009, cap. 5.
 - supuestos, estructura e interpretación
 - DiD con interacciones
- *Práctica*:
 - corriendo modelos DiD
 - visualización de los resultados
- Consigna para la *Tarea #8*.

Sesión 13 (Noviembre/30). Replicación (3)

- Entrega y revisión de la *Tarea #8*.
- *Lectura*: texto a replicar: Edmund J. Malesky, Cuong Viet Nguyen, and Anh Tran. The Impact of Recentralization on Public Services: A Difference-in-Differences Analysis of the Abolition of Elected Councils in Vietnam. *American Political Science Review*, 108(1):144–168, 2014.
- *Replicación*: datos de panel + diferencia-en-diferencias.

Sesión 14 (Diciembre/7). Repaso

- Entrega y revisión de la *Replicación #3*.
- Repaso de temas y discusión de trabajos finales. Asistencia optativa.