

Universidad Católica Andrés Bello
Escuela de Economía
Introducción a la Inferencia Causal

Profesores:

José Morales-Arilla, Candidato a PhD en Políticas Públicas en la Universidad de Harvard
Carlos Daboín, Consultor del Workforce of the Future Initiative, The Brookings Institution

Objetivo:

El propósito de este curso es ofrecer a los estudiantes conocimientos conceptuales y prácticos para realizar análisis de efectos causales. El curso está abierto a estudiantes de la escuela de economía y a estudiantes de otras escuelas y universidades que estén interesados en el material y cuenten con los requisitos previos.

Requisitos:

Contar con entrenamiento previo en estadística, álgebra lineal y análisis de regresión es fundamental. Cualquier experiencia en el manejo de datos con herramientas computacionales (R, Python, Stata, Excel, etc.) es considerada conveniente.

Dinámica:

El curso se divide en cinco bloques con tres clases en cada uno. Se asignan trabajos prácticos al final de cada bloque, en los cuales el estudiante usará datos reales para abordar los problemas mencionados en el mismo. Dichas asignaciones tendrán un componente práctico y computacional importante, requiriendo el uso de softwares de análisis estadístico y una importante cantidad de trabajo fuera del aula.

Software:

Las asignaciones y el material del curso serán presentados en lenguaje R, por lo que se insta a los estudiantes a familiarizarse con dicho programa. Se admite que el estudiante use el software de su preferencia al realizar sus trabajos prácticos, pero los profesores no se comprometen a brindar apoyo de programación en otros softwares.

Evaluación:

- Trabajo final y presentación: 30% (15% c/u)
 - Los trabajos serán en grupo de hasta 5 personas.
- Asignaciones: 60% (15% c/u).
 - Los lineamientos de cada asignación serán brindados tras la clase número 3 del bloque.
- Participación: 10%
 - Se considerará participación tanto contribuciones en clase como en el foro de discusión.

Código de conducta:

Si bien invitamos la colaboración entre estudiantes, cada estudiante debe entregar respuestas individuales a las asignaciones. En el foro de discusión los estudiantes podrán interactuar y hacer preguntas basadas en código, pero está prohibido compartir códigos enteros para responder a preguntas o a asignaciones.

Temario:

Semana 1: Introducción y Evidencia Experimental

Clase #1 - FECHA - Introducción al curso

- Introducción y logística
- Revisión de estadística y econometría
- Introducción a R

Clase #2 - FECHA - Introducción a la inferencia causal

“Correlación no implica causalidad”

- Diagramas de causalidad
- Esquema de resultados potenciales
- Evidencia experimental vs. observacional
 - Sesgos de variable omitida: Fumar, edad y cáncer.
 - Endogeneidad y causalidad inversa: Crimen y actividad económica.

Clase #3 - FECHA - Evidencia Experimental

“Los experimentos garantizan asignación aleatoria”

- Experimentos de asignación aleatoria (RCTs).
 - Mullainathan - Racismo en el mercado laboral.
- Experimentos naturales.
 - Michelopolous y Pappaionau - Scramble for Africa
- Evaluación de evidencia experimental
 - Cámaras en escuelas en la India.
 - Anti-parasíticos en Kenya.

Tarea #1 - Lineamientos tras Clase #3, a entregarse antes del inicio de la Clase #4

- Evaluación de impacto #1
- Evaluación de impacto #2
- Preguntas de investigación:
 - Plantea 2 preguntas de investigación desde la perspectiva de causalidad.
 - Presentar un diagrama de causalidad.
 - Discutir desde la perspectiva del esquema de resultados potenciales.
 - Trata de enfocarte en temas que piensas te gustaría o podrías trabajar en tu tesis de grado (no tienes que profundizar, pero piensa en la viabilidad del estudio desde la perspectiva de datos),

Semana 2: Evidencia cuasi-experimental en datos de sección cruzada

Clase #4 - FECHA - Matching:

“Vincular a gemelos no tratados puede ayudar a detectar efectos causales”

- Evidencia cuasi-experimental y supuestos de identificación
- Lógica de “matching” para estimación de efectos causales
- Subclasificación
- “Matching” exacto
- “Matching” aproximado

Clase #5 - FECHA - Regresión de discontinuidad:

“Cambios discretos en la probabilidad de ser tratado ayudan a

- La lógica de la regresión de discontinuidad
- Estimación con regresión de discontinuidad
- Retos a los supuestos de identificación de la regresión de discontinuidad

Clase #6 - FECHA - Variables instrumentales

“Un buen instrumento solo afecta a la variable de resultado a través del tratamiento”

- Historia, intuición y supuestos de las variables instrumentales.
- Efectos homogéneos y variables instrumentales.
- El problema de los instrumentos débiles.
- Efectos heterogéneos.
- Aplicaciones
 - RCTs con incumplimiento del tratamiento (Loterías)
 - Microfinanzas
 - Promedio en otras unidades (Efectos fijos por juez)
 - Efectos de sentencias más largas
 - Instrumentos de Bartik (Cambio y Porción)
 - Efectos locales de la competencia china

Tarea #2: Lineamientos tras clase #6, a entregarse antes del inicio de la Clase #7.

- Ejercicio de matching
- Ejercicio de regresión de discontinuidad
- Ejercicio de variables instrumentales
- ¿Cómo evaluarías tu pregunta de investigación?
 - ¿Cuál sería el ideal experimental en tu pregunta?
 - ¿Cuáles son las dificultades de implementar el ideal experimental?
 - ¿Cómo podría abordarse la pregunta de forma cuasi-experimental?

Semana 3: Evidencia cuasi-experimental en datos de panel

Clase #7 - FECHA - Datos de panel

“Los cambios en el tiempo pueden ayudar a identificar efectos causales”

- Introducción a datos en formato panel.
- MCO agrupados
- Estimadores de efectos fijos
- Ejercicios prácticos

Clase #8 - FECHA - Diferencia en diferencias

“Para que el cambio asociado con en el tratamiento se entienda como causal, este no debe explicar los cambios ocurridos antes del tratamiento”

- Introducción a diferencia en diferencias.
- Estimación e inferencia
- Evidencia de tendencias paralelas previo al tratamiento
- Problemas de diferencia en diferencias con tratamientos en tiempos distintos

Clase #9 - FECHA - Controles sintéticos

“Los controles sintéticos mezclan matching con diferencia en diferencias para hacer casos de estudio”

- Introducción a los controles sintéticos
- Casos de estudio comparativos con controles sintéticos
- Ejercicios y discusión

Tarea #3 - Lineamientos tras clase #9, a entregarse antes del inicio de la clase #10

- Ejercicio de regresión de panel
- Ejercicio de diferencia en diferencias
- Ejercicio de control sintético
- ¿Qué datos necesitarías para evaluar tu pregunta de investigación?
 - Discute en más detalle el contexto de tu estudio.
 - ¿Qué datos necesitarías para medir el resultado de interés?
 - ¿Qué datos necesitarías para medir el tratamiento de interés?
 - ¿Qué datos necesitarías como controles de interés?

Semana 4: Introducción a Machine Learning y Big Data

Clase #10 - FECHA - Principios de Machine Learning para predicción

“Si te interesa la predicción, tienes que predecir fuera de muestra”

- Machine Learning “supervisado”
 - El trade-off entre varianza en las predicciones y sesgo en los estimados
 - “Overfitting”: El sesgo entre flexibilidad y precisión.
 - Problemas de predicción.
 - Problemas de clasificación.
 - Modelos lineales
 - Modelos no-lineales
- Machine Learning en ejercicios de inferencia causal
 - Preguntas muy interesantes solo sobre predicción y no causalidad
 - Limpieza de datos con observaciones faltantes
 - Selección de modelo e identificación de controles

Clase #11 - FECHA - Todo es data (1)

- Machine Learning “no-supervisado”
 - Agrupación en la data (Clustering)
 - Detección de comunidades
- Análisis y visualización de redes
- Análisis y visualización de data geoespacial

Clase #12 - FECHA - Todo es data (2)

- Extracción de data desde el internet
 - Web scraping
 - APIs - Data banks
- Análisis y visualización de data en formato de texto
- Visualización de datos estática e interactiva
 - Ggplot2
 - Shiny

Tarea 4 - Lineamientos tras clase #12, a entregarse antes de la clase #13

- Problema sencillo de Machine Learning
- Análisis de cluster
- Visualización geoespacial de datos
- Visualización de redes de datos
- ¿Qué análisis estadístico harías para responder tu pregunta causal?
 - Escribe la regresión
 - Explica tu supuesto de identificación
 - ¿Cuáles son las principales amenazas a tu supuesto de investigación?
 - ¿Cómo podrías evaluar la validez de tus supuestos?

Semana 5: Discusión amplia de la literatura empírica y presentaciones de los estudiantes

Clase #13 - FECHA - Revisión de literatura

- Estudios experimentales
- Estudios en sección cruzada
- Estudios de panel
- Estudios de Machine Learning

Clase #14 - FECHA - Discusión especial con invitados

- Discusión especial

Trabajo final y presentación - Lineamientos tras clase #9, a entregarse antes de la clase #15

- Diapositivas y documento con proyecto de investigación.

Clase #15 - FECHA - Presentaciones de los estudiantes.

Referencias:

Cunningham, Scott. 2021. Causal Inference The Mixtape. New Jersey: Yale University Press.

Cunningham, S. (2021). Causal Inference The Mixtape. Yale University Press.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning -- with Applications in R (Vol. 103). New York: Springer. ISBN: 978-1-4614-7137-0

Angrist, J., Pischke, J.-S. (2015). Mastering 'Metrics: The path from cause to effect. Princeton University Press.

Angrist, J. D., Pischke, J.-S. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press. ISBN: 0691120358

Wickham, H., Golemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media. ISBN: 1491910399