

1. Información del Curso

Profesor: Mauricio Romero
Horario: Lunes 11:30 AM - 1:00 PM
Salón: RH-313
E-mail: mtromero@itam.mx
Horario de Atención: Por Zoom con cita previa

2. Objetivos del Curso

Este seminario es una introducción práctica al trabajo con datos reales. El objetivo central es que el alumno aprenda a “ensuciarse las manos” con los datos: desde cómo descargarlos y limpiarlos, hasta cómo visualizarlos correctamente para contar una historia. Se pondrá énfasis en entender por qué “correlación no es causalidad” a través de la intuición detrás del concepto de endogeneidad, preparando el terreno para cursos futuros más avanzados.

Al finalizar el curso, los estudiantes deberían ser capaces de:

- **Gestionar bases de datos:** Limpiar, fusionar datos (e.g., merge/append) y ordenar (e.g., reshape) fuentes reales (INEGI, Banco Mundial, etc.).
- **Visualizar información:** Crear gráficos que comuniquen hallazgos de manera clara, estética y honesta.
- **Entender el problema de identificación:** Explicar intuitivamente qué es el sesgo de selección/endogeneidad y por qué es difícil medir el impacto causal de una política pública.

3. Contenido y Cronograma

Nota: Este cronograma es tentativo y se puede ir ajustando a medida que avanza el semestre.

12 de Enero	Introducción: ¿De dónde vienen los datos? Encuestas (Censos, ENIGH) vs. Registros Administrativos vs. Big Data.
19 de Enero	Introducción a R: Instalación, interfaz y primeros pasos.
26 de Enero	Usando datos en R (Parte I): Importación y limpieza básica.
09 de Febrero	Usando datos en R (Parte II): Manipulación de datos.
16 de Febrero	Conceptos básicos de programación: Loops, condicionales y funciones.
23 de Febrero	Visualización de datos: Principios gráficos y ‘ggplot2’.

02 de Marzo	Parcial 1
09 de Marzo	Correlación vs. Causalidad (Parte I): Intuición y ejemplos.
23 de Marzo	Correlación vs. Causalidad (Parte II): El problema del contrafactual.
06 de Abril	Experimentos (Parte I): El “Gold Standard”.
13 de Abril	Experimentos (Parte II): Aplicaciones y límites.
20 de Abril	Parcial 2
27 de Abril	Presentaciones Proyecto Final I
04 de Mayo	Presentaciones Proyecto Final II
11 de Mayo	Presentaciones Proyecto Final III

Fechas Importantes:

- **Parcial 1:** 02 de Marzo
- **Parcial 2:** 20 de Abril
- **Último día para darse de baja:** 29 de Abril

4. Tareas y Evaluación

La única manera de aprender los contenidos del curso es aplicando lo aprendido. El curso tendrá talleres regulares que incluirán ejercicios teóricos y prácticos. Las entregas deben realizarse de forma individual, aunque se permite discutir ideas con compañeros. El código debe ser claro, replicable y escrito desde cero.

Componente	Porcentaje
Talleres (Tareas prácticas)	25 %
Parcial 1	25 %
Parcial 2	25 %
Trabajo Final (Reporte + Código replicable)	25 %

Nota: No hay examen final durante la semana de exámenes.

5. Asistencia

La asistencia no es obligatoria pero sí **altamente recomendada**. Las clases están diseñadas para tener discusiones activas, resolver dudas de código y trabajar con datos/simulaciones en tiempo real.

6. Uso de Inteligencia Artificial

Pueden usar herramientas de IA (como ChatGPT o Copilot) para ayudar en la codificación o depuración de código. Sin embargo:

- Deben entender línea por línea el código que entregan.
- No pueden copiar sin comprender.
- Son totalmente responsables de la exactitud del contenido que entreguen.

“Usar IA para hacer tus tareas es como enviar un robot al gimnasio para hacer ejercicio por ti. No te vas a poner más fuerte.” Hagan el trabajo. Se aprende haciendo.

7. Bibliografía

Los siguientes textos sirven como referencia general para el curso:

- Huntington-Klein, Nick. *The Effect: An Introduction to Research Design and Causality*. <https://theeffectbook.net/index.html>
- Cunningham, Scott. *Causal Inference: The Mixtape*. <https://www.scunning.com/mixtape.html>
- Angrist, Joshua and Pischke, Jörn-Steffen. *Mostly Harmless Econometrics*. Princeton University Press.
- Angrist, Joshua and Pischke, Jörn-Steffen. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

8. Recursos y Herramientas Computacionales

A continuación se listan recursos útiles para el desarrollo de la clase y su futuro profesional.

R y Programación

- **Instalación:** Primero instale la última versión de [R-Project.org](https://www.r-project.org/) y posteriormente [RStudio.com](https://rstudio.com/).
- **Oferta para Estudiantes:** Se recomienda solicitar el [GitHub Student Developer Pack](#) para acceso gratuito a herramientas profesionales.
- **Guías y tutoriales de R:**
 - [swirl](#): Aprende R dentro de R.
 - [Base R cheat sheet](#) y [R Studio Resources](#).
 - [R-bloggers](#): Blog con noticias y tutoriales.
 - [Kelsey Moty's R Workshops](#) (NYU CDSC Lab).

- **Ciencia de Datos y Economía en R:**

- Grant McDermott: [Data Science for Economist class](#) (Repo: ds4e).
- Grant McDermott: [Efficient Simulation in R](#).
- Ljubica “LJ” Ristovska: [Language-Agnostic Guide to Programming for Economists](#).
- Jesse M. Shapiro and Matthew Gentzkow: [Code and Data for the Social Sciences](#).
- DIME (World Bank): [R-training course](#) (GitHub).

- **Herramientas Auxiliares:**

- [TablesGenerator](#): Para convertir tablas de Excel a $\text{\LaTeX}$.
- [Layout Parser](#): Librería de *deep-learning* para procesar imágenes de documentos a escala.

Visualización y Comunicación

- Kieran Healy: [Visualization: A Practical Introduction](#).
- Jonathan A. Schwabish: [An Economist’s Guide to Visualizing Data](#)
- Esquemas de color:
 - [Paquete ‘viridis’ en R](#) (Accesible para daltónicos - Recomendado).
 - [MulinBlog Palette](#)
- J-PAL: [Data visualization resources](#).
- Alex Coppock: [Visualizations for RCTs](#)
- Monica Alexander: [Blog y Recursos](#)
- **Hablar en Público:**
 - Jesse M. Shapiro: [How to Give an Applied Micro Talk](#).
 - Rachael Meager: [Public Speaking for Academic Economists](#).

Recursos de Stata (Referencia)

Aunque el curso prioriza R, estos recursos son valiosos para referencia general:

- [IPA’s Stata training](#).
- [UCLA’s Stata resources](#).
- [Stata introduction by German Rodriguez](#).
- [Stata introduction by Oscar Torres-Reyna](#).
- J-PAL: [Data cleaning in Stata](#).