

Métodos Cuantitativos I - ICE3201

Statistics: "science dealing with data about the condition of a state or community"

Gottfried Aschenwall, 1770

Universidad de O'Higgins
Instituto de Ciencias Sociales
Rancagua, Chile

Última actualización: August 5, 2020.
Descarga la última versión [aquí](#).

Aspectos Logísticos

Profesor: Héctor Bahamonde, PhD.
e: hector.bahamonde@uoh.cl
w: www.HectorBahamonde.com
Zoom ID: 951-326-1038.
Office Hours (Zoom): Toma una hora [aquí](#).

Hora de cátedra: Lunes (12-12.40) y jueves (14.30-15.10)
Lugar de cátedra: Zoom (no hay clases presenciales este semestre).
Link Zoom Lunes: <https://us02web.zoom.us/j/82818093668>, o meeting ID 828-1809-3668.
Link Zoom Jueves: <https://us02web.zoom.us/j/82331757482>, o meeting ID 823-3175-7482.

Acceso a materiales del curso: [uCampus](#).

Ayudante de cátedra (TA): Gonzalo Barría.
e: gonzalo.barria@uoh.cl
Zoom ID: 988-891-7227.
TA Bio: Gonzalo Barría es Cientista Político (PUC) y Magíster en Ciencia Política (PUC).
Hora de ayudantía: **PENDIENTE**.
Lugar de ayudantía: Zoom (no hay ayudantías presenciales este semestre).
Link Zoom ayudantía: **PENDIENTE**, o meeting ID PENDIENTE.

Carrera: Ingeniería Comercial.
Eje de Formación: Métodos Cuantitativos.
Semestre/Año: Quinto Semestre/2020.
Pre-requisitos: Métodos Matemáticos Avanzados y Teoría Estadística.
SCT: 6.
Horas semanales: Cátedra (30-40 minutos vía Zoom), Ayudantía (30-40 minutos vía Zoom).
Semanas: 15.

Motivación: ¿Por qué tomar este curso?

¿Qué efecto tiene la educación sobre los ingresos? ¿Cómo podemos evaluar los efectos de una reforma educacional? ¿La legalización de las drogas aumenta su consumo? ¿Qué candidato/a ganaría la elección presidencial si ésta fuera mañana?

Las entidades públicas guían sus decisiones estratégicas en base a información cuantificable, i.e. datos. Esto ha tomado incluso más importancia en la actualidad, donde ha habido una digitalización de los datos sociales. Es fundamental que los científicos sociales en general sepan cómo usar estos datos. Aún más, el quehacer social en general, está constantemente produciendo datos. Cada vez que usas *Twitter*, pides un *Uber*, envías un e-mail, votas, respondes una encuesta, estás produciendo datos sociales. Piensa en lo siguiente: si bien es cierto que hace unos diez años atrás *faltaban* datos, hoy en día los datos *sobran*. El desafío actual consiste en saber cómo analizarlos correctamente, y así ayudar a los tomadores de decisiones. Esto es importante. Mañana tu podrías ser un/a analista en una de las decenas de Departamentos de Estudios repartidas en la administración del Estado. **Este curso te prepara para ese mundo** (incluyendo el mundo de la consultoría).

Aunque lo que aprenderemos es altamente numérico y matemático, no te confundas. Estos métodos no son infalibles, y no nos contarán “la verdad” (si es que algo así existiera). Aún necesitas ser muy crítico(a). Como verás, **la estadística inferencial (que es el objeto de este curso) es un arte, no una ciencia**. Los números nos sugerirán ciertas ideas, pero aun así nuestro trabajo será *interpretar* estos resultados. No seas obediente. Se crítico/a y auto-crítico/a. Sospecha de tus propios resultados y el de los demás. Mal que mal, estaremos haciendo **inferencias** (no *certezas*) estadísticas. Como veremos, el fantasma de este semestre se llamará *incertidumbre*.

Este curso considera un énfasis especial en la *causalidad*. La *inferencia causal* ha llegado para quedarse en las ciencias sociales. *¿Bajo qué condiciones podemos decir que X causa Y?* Más que una cuestión matemática, la causalidad toca en muchos aspectos la filosofía de las ciencias. Este semestre aprenderemos qué relación tiene la experimentación con la causalidad, cómo podemos hacer experimentos en ciencias sociales, y cómo podemos emular un experimento (usando ciertos métodos estadísticos) cuando no podemos ni debemos hacer uno.

Honestamente, espero que este curso cautive tu atención, y simiente tu curiosidad intelectual, sobre todo, mostrándote que nuestro objeto de estudio (la sociedad) es apasionante.

Bienvenid@s!

Propósito Formativo

El objetivo de este curso es introducir al/la alumno/a a los métodos econométricos básicos para el análisis de datos. El curso avanza progresivamente en distintos tópicos en regresión lineal y métodos no lineales. La principal característica es la introducción a modelos de regresión lineal para en cursos más avanzados estudiar otro tipo de estimaciones.

Objetivos Generales del Curso

El gran objetivo de este curso, es poder generar en la/el estudiante la capacidad de razonamiento crítico, desde un punto de vista empírico.

El lenguaje que aprenderemos este semestre será R, el lenguaje de programación más usado en las ciencias sociales. Esto tiene varias ventajas. R es gratis y corre en todas las plataformas disponibles. Segundo, es

un lenguaje orientado a “objetos”. Esto significa—tercero—que fuerza al/la estudiante a realmente pensar en el proceso matemático/estadístico detrás del análisis que se está haciendo. Al contrario de otros *softwares* estadísticos como SPSS y Stata, donde el/la usuario(a) simplemente aprieta botones sin saber lo que ocurre realmente, R necesita que le digamos exactamente qué hacer. Y eso es lo que aprenderemos este semestre. Cuarto, si sabes R, te será absolutamente fácil aprender Stata (o SPSS).

A pesar de estas ventajas, los economistas siguen ocupando Stata. Es por esto que también cubriremos las funciones básicas en Stata. Dado que Stata es pagado y que está disponible solamente en nuestro laboratorio de computación UOH, se contempla la posibilidad de hacer un taller corto, voluntario y sin nota el segundo semestre 2020.

Este curso está dividido en cuatro grandes unidades. Cada unidad tiene sus diferentes evaluaciones.

1. Funciones básicas en R.
2. Estadística descriptiva en R.
3. Introducción a modelos lineales en R.
4. Inferencia causal en R.

Objetivos Específicos del Curso

1. Lograr establecer una pregunta económica y un método de identificación que permita verificar la hipótesis de forma causal.
 2. Poder *testear* hipótesis y tener las herramientas para analizar políticas de forma crítica.
 3. Entender las limitaciones de los trabajos empíricos y los *trade offs* existentes al establecer supuestos.
- Se espera que los estudiantes hagan sus respectivas lecturas *antes* de cada clase para poder participar en el debate crítico que haremos en cada una de ellas. También se espera que los/las estudiantes hagan los ejercicios prácticos clase a clase.

Instalación de R

Primero, instala R desde el [sitio Web](#) oficial. Click en “CRAN” (extremo superior izquierdo). Selecciona cualquier *mirror*. Por ejemplo, bájalo desde el *o-Cloud*. Después, baja la interfaz más utilizada, llamada R-Studio. Para esto, anda al [sitio Web](#) oficial, después *Download R-Studio, FREE*, selecciona la versión que sea compatible con tu sistema operativo (Windows, Mac, Ubuntu).

Política sobre Trabajo Cooperativo

Este es un curso práctico. Espero que pases muchas horas frente a la pantalla de tu computador personal. **Yo recomiendo el trabajo cooperativo.** Es saludable que consultes con tus compañeros/as de curso, y que traten, en la medida de lo posible, de encontrar las soluciones en conjunto. Sin embargo, salvo por el examen final y la presentación final (más sobre esto abajo), todos los trabajos (y sus evaluaciones) serán individuales.

Evaluaciones

1. Lecturas, Participación y Pop Quizzes: 15%.

El TA y yo asumiremos durante todo el semestre que has leído. Nosotros empleamos un método de clases interactivo, pero este método necesita de tu participación activa en clases.

Para asegurarnos de que estés haciendo las lecturas, habrán una serie de *pop quizzes* (“pruebas sorpresa”) tanto en cátedra como ayudantía. Estos controles serán cortos (5-10 minutos), y apuntan a medir si leyeron; o sabes, o no sabes. Estas pruebas se aplicarán completamente al azar, en cualquier momento de la clase, y sin previo aviso. En general, las preguntas serán acerca de un concepto clave, y cuya respuesta correcta será una línea (o dos, como máximo).

2. “Ejercicios Prácticos Chicos” en clases (vía Zoom): 5% cada uno, 25% en total.

- (a) Funciones básicas en R: #1, #2. Total= 10%.
- (b) Estadística descriptiva en R: #3. Total= 5%.
- (c) Introducción a modelos lineales en R: #4, #5. Total= 10%.

Tendrás 3 días corridos para entregar estos trabajos via uCampus.

- ☞ Es importante que estas líneas corran bien: el usuario (yo) tiene que ser capaz de ver cómo R ejecuta cada línea, sin estancarse.
- ☞ Es importante que vayas guiando al usuario (yo) sobre tu raciocinio. Asegúrate de comentar (usando el símbolo #).

3. “Tareas grandes” para la casa, una por unidad: 10% cada uno, 40% en total. Se borra la nota más baja (30% en total final).

- (a) Funciones básicas en R: 10%.
- (b) Estadística descriptiva en R: 10%.
- (c) Introducción a modelos lineales en R: 10%.
- (d) Introducción a la inferencia causal en R: 10%.

En estos ejercicios deberás resolver un problema práctico. Según lo estipula el programa, recibirás una base de datos, y una serie de preguntas de carácter aplicado. El producto (i.e. lo que tienes que entregar), será un *script* de R. Un *script* es un texto que contiene líneas de programación (de R), que al ser ejecutadas, me llevarán a tu respuesta. El plazo para entregar el *script* de una semana una vez recibidas las instrucciones. Se entrega vía uCampus.

☞ Aunque no es necesario, sí puedes ocupar recursos externos, como Internet.

- ☞ Es importante que estas líneas corran bien: el usuario (yo) tiene que ser capaz de ver como R ejecuta cada línea, sin “estancarse”.
- ☞ Es importante que vayas guiando al usuario (yo) sobre tu raciocinio. Asegúrate de comentar (usando el símbolo #).

4. Un examen final obligatorio/no-eximible (20%) y una presentación final (10%, vía Zoom): 30% en total.

En este curso, la actividad final es un examen final (20%) que tiene formato de trabajo grupal. Usando una base de datos, tú y tu grupo deberán responder una serie de preguntas. El producto final

(i.e. lo que debes entregar) consiste en un *script* de R. La nota es grupal (i.e. todo el grupo recibirá la misma nota). **Los grupos serán de 2 personas.**

El paper (*script*) se puede entregar antes, pero una vez cerrado el plazo, no se recibirán trabajos. Los *scripts* que se entreguen tarde o vía *email* tendrán un 1 (sin opción a reclamo). **No hay excepciones.**

En un formato muy parecido a una conferencia académica (virtual, no presencial), tendrás (junto a tu grupo) que presentar los principales hallazgos (10%). Todos/as presentan. Cada presentación debe durar no menos de 15 minutos, pero nunca más de 20 minutos. Las presentaciones se realizarán virtualmente (i.e. vía Zoom) el último día de clases. Tendrás que ocupar *slides* ("Power Point"). Para tales efectos, tendrás que compartir pantalla desde tu casa, y hacer tu presentación de esa manera.

Les recomiendo "verme" (vía Zoom) en [mis office hours](#) antes del plazo de entrega. Si quieres, [envíame un email](#) con tu borrador, y yo te devolveré comentarios. Vélo como una pre-corrección. Esto es voluntario. También puedes contactar al/la TA. **No se procesarán preguntas durante fines de semana, y/o festivos.**

En resumen:

| | Porcentaje | Porcentaje Acumulado |
|--|------------|----------------------|
| Participación y <i>pop-quizzes</i> (cátedra y ayudantía) | 15% | 15% |
| Evaluación práctica "chica" (taller en clases, vía Zoom): #1 | 5% | 20% |
| Evaluación práctica "chica" (taller en clases, vía Zoom): #2 | 5% | 25% |
| Evaluación práctica "chica" (taller en clases, vía Zoom): #3 | 5% | 30% |
| Evaluación práctica "chica" (taller en clases, vía Zoom): #4 | 5% | 35% |
| Evaluación práctica "chica" (taller en clases, vía Zoom): #5 | 5% | 40% |
| Tarea práctica "grande" para la casa: #1 | 10% | 50% |
| Tarea práctica "grande" para la casa: #2 | 10% | 60% |
| Tarea práctica "grande" para la casa: #3 | 10% | 70% |
| Tarea práctica "grande" para la casa: #4 | 10% | 80-10=70% |
| "Examen" final grupal | 20% | 90% |
| Presentación grupal | 10% | 100% |

Ayudantía

Cada semana te reunirás con tu ayudante ("TA") vía Zoom. Ahí tendrás otra oportunidad para ejercitar y seguir profundizando otras temáticas pendientes. En esta oportunidad, también se revisarán aspectos más formales de las humanidades y las ciencias sociales.

Calendario

1. Funciones básicas en R

• Clase #1

- Introducciones: programa de curso, requerimientos, expectativas, etc.
- *Qué es R?* Instalación de R y RStudio.
- *Qué es Stata?*
- **Lecturas:**

- ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: Cap. 1.
- ◇ Urdinez y Cruz, 2019. *AnalizaR Datos Políticos*: Cap. 2.

- **Clase #2**

- Funciones básicas: promedio, `help()`, operadores, tipos de objetos (*character*, *arrays*, fechas, listas, *dataframes*).
- Cargando bases de datos (I): formatos, etiquetas, tipos de variables, descripción básica.
- **Lecturas:**
 - ◇ Urdinez y Cruz, 2019. *AnalizaR Datos Políticos*: Cap. 5.

- **Clase #3**

- ✉ Evaluación práctica “chica” (taller en clases, vía Zoom): #1.

- **Clase #4**

- Cargando bases de datos (II): transformaciones, creación de nuevas variables.
- Manipulando bases de datos: generación de matrices y *dataframes*, `merge`, `append`. Logs.

- **Clase #5**

- Visualización de datos (I): *bar plots*, *scatter plots*, histogramas, *time series plots*.
- **Lecturas:**
 - ◇ Urdinez y Cruz, 2019. *AnalizaR Datos Políticos*: Cap. 4.

- **Clase #6**

- Visualización de datos (II): *plots* más complejos (por categorías), mapas.
- **Lecturas:**
 - ◇ Urdinez y Cruz, 2019. *AnalizaR Datos Políticos*: Cap. 15.

- **Clase #7**

- ✉ Evaluación práctica “chica” (taller en clases, vía Zoom) #2.
- ★ Entrega del temario para la tarea práctica “grande”: #1.

2. Estadística descriptiva en R

- **Clase #8**

- Estadística descriptiva (I): Teoría de probabilidades: distribuciones, varianza.

- **Clase #9**

- Estadística descriptiva (II): binomial, normal, otras; simulación.

- **Clase #10**

- ✉ Evaluación práctica “chica” (taller en clases, vía Zoom): #3.
- ★ Entrega del temario para la tarea práctica “grande”: #2.

3. Introducción a modelos lineales en R

- **Clase #11**

- Introducción a modelos lineales: *Que es OLS?*
- **Lecturas:**

- ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: 2.1 y 2.2.

- **Clase #12**

- La mecánica detrás del OLS (I): matrices “a mano”.
- **Lecturas:**
 - ◇ Krishnan Namboodiri, 1984. *Matrix Algebra, an Introduction*. Sage: Caps. 1 y 2.

- **Clase #13**

- La mecánica detrás del OLS (II): matrices en R.

- **Clase #14**

- Coeficientes.
- **Lecturas:**
 - ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: Caps. 3.1 y 3.2.

■ **Evaluación práctica “chica” #2, #3 y #4.** Entrega viernes 26, 5pm uCampus. Evaluaciones grandes #1 y #2 quedan para la semana después del *break*.

- **Clase #15**

- Error, residual y ϵ_i .

- **Clase #16**

- Intervalos de confianza.
- **Lecturas:**
 - ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: Cap. 4.3.

- **Clase #17**

- Test de hipótesis (*t test*), errores Tipo I y II, significación estadística (*p-values*).
- **Lecturas:**
 - ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: Cap. 4.2.

- **Clase #18**

- Propiedades numéricas del OLS, Gauss-Markov, sesgo de variable omitida.
- **Lecturas:**
 - ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: pp. 89-94, 102-104.

- **Clase #19**

- *Goodness of fit*, “coeficiente de determinación” (r^2), predicción.
- **Lecturas:**
 - ◇ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: pp. 40-41, Cap. 6.3.
 - ◇ Gary King, 1986. *How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science*. American Journal of Political Science(30): 666-687.

- **Clase #20**

- Problemas y *post-estimation*: multicolinealidad perfecta, heteroskedasticidad, no linealidad, *outliers*, no normalidad de residuos, auto-correlación.
- **Lecturas:**
 - ◊ Jeffrey Wooldridge, 2010. *Introducción a la Econometría. Un Enfoque Moderno*. Cengage Learning: Caps. 8 y 9.5.

4. Inferencia causal en R

- **Clase #21**

- El “gold standard”. Correlación y causación. Modelos lineales y causalidad.
- **Lecturas:**
 - ◊ David Freedman. 1991. “*Statistical Models and Shoe Leather.*” *Sociological Methodology*, 21(1): 291-313.
 - ◊ David Freedman. 1999. “*From Association to Causation: Some Remarks on the History of Statistics.*” *Statistical Science*, 14(3): 243-258.
 - ◊ Joshua Angrist and Jorn-Steffen Pischke, 2008. *Mostly Harmless Econometrics*. Princeton University Press: section 3.2.

- **Clase #22**

- Variables instrumentales y *two-stage least squares*.
- **Lecturas:**
 - ◊ Joshua Angrist and Jorn-Steffen Pischke, 2008. *Mostly Harmless Econometrics*. Princeton University Press: section 4.1-4.2.

- **Clase #23**

- Regression discontinuity designs.
 - **Lecturas:**
 - ◊ Joshua Angrist and Jorn-Steffen Pischke, 2008. *Mostly Harmless Econometrics*. Princeton University Press: chapter 6.
- ★ Entrega del temario para la tarea práctica “grande”: #4.

- **Clase #24**

- Incorporando el elemento *tiempo*: fixed effects, differences-in-differences.
- **Lecturas:**
 - ◊ Joshua Angrist and Jorn-Steffen Pischke, 2008. *Mostly Harmless Econometrics*. Princeton University Press: chapter 5.

- **Clase #25**

- Asignación aleatoria de bases de datos, y entrega de preguntas. Conformación de grupos de trabajo.
- **Laboratorio #1:** Trabajo en grupo, y resolución de preguntas.

- **Clase #27:**

- **Presentaciones finales.**

References

Namboodiri, Krishnan. 1984. *Matrix Algebra: An Introduction*, 1–99. Sage.

Urdinez, Francisco, and Andrés Cruz. 2019. *AnalizaR Datos Políticos*. Edited by Francisco Urdinez and Andrés Cruz. <https://arcruzo.github.io/libroadp/>.

Wooldridge, Jeffrey. 2010. *Introducción a la Econometría. Un Enfoque Moderno*. 4th. Cengage Learning.