

Métodos Cuantitativos II - ICE3202

"There is therefore an absolute measure of probability [...] There is no such absolute measure of likelihood"

Sir Ronald Fisher, 1922

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

Última actualización: November 19, 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: Martes: 10:15—11:45; Jueves: 14:30—16:00 hrs.
Lugar de cátedra: Zoom (no hay clases presenciales este semestre).

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

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

Carrera: Ingeniería Comercial.
Eje de Formación: Métodos Cuantitativos.
Semestre/Año: Sexto Semestre/2020.
Pre-requisitos: Metodos Cuanti I.
SCT: 6.
Horas semanales: Cátedra (45-60 minutos vía Zoom), Ayudantía (45-60 minutos vía Zoom).
Semanas: 12.

Motivación: ¿Por qué tomar este curso?

Muchas de las preguntas que nos hacemos como científicos sociales tienen que ver con variables “especiales”. Los métodos lineales como el OLS suponen variables continuas (por ej., GDP, ingresos, desigual-

dad). Sin embargo, existen otras variables igualmente interesantes como participación electoral (si la persona votó o no). Esta variable, por ejemplo, es binaria. Existen otras que son categóricas. Imagínate estás estudiando el tipo de transporte usado para llegar a tu lugar de trabajo: bus, auto, taxi. Cómo puedes estimar un modelo cuya variable dependiente no tiene un orden? Es imposible ordenar las categorías “bus”, “auto” y “taxi” en una escala. En estos y otros casos es necesario usar otro método de estimación diferente al que habíamos visto antes. Este semestre aprenderemos como estimar modelos generalizados GLMs (*generalized linear models*) vía MLE (*maximum likelihood estimation*).

Si la tarea de antes (OLS) era **minimizar** los errores cuadrados, la tarea de este semestre será **maximizar** el *likelihood*.

Qué diferencias existen entre la “probabilidad” y el “*likelihood*”? Qué diferencias matemáticas existen entre OLS y MLE? Cómo seleccionar el modelo adecuado para cada tipo de variable?

Este semestre, además, prestaremos especial atención a inferencia causal.

Bienvenid@s al mundo de los modelos generalizados y estimados vía maximum likelihood estimation!

Propósito Formativo

El objetivo de este curso es profundizar los conocimientos introducidos por Métodos Cuantitativos I, particularmente, inferencia causal. Este curso proporciona los fundamentos teóricos y prácticos de la econometría tradicional. El principal objetivo es profundizar los conceptos y construir un conocimiento sólido sobre los fundamentos inducidos por cursos anteriores. Además, el enfoque es entregar otro tipo de estimaciones que permitan solucionar los problemas estadísticos prácticos.

Se espera que el alumno logre:

1. Logre establecer una pregunta económica y un método de identificación que permita verificar la hipótesis de forma causal.
2. Poder *testear* hipótesis y tengan las herramientas para analizar políticas de forma crítica.
3. Entender las limitaciones de los trabajos empíricos y los *trade offs* existentes al establecer supuestos.
4. Manejar distintas bases de datos con un propósito de investigación.
5. Pleno manejo en algún software estadístico.

Software

En este curso usaremos principalmente R. Sin embargo, y debido a que Stata sigue siendo uno de los programas usados en la disciplina, también lo abordaremos.

- 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). Es importante saber si tu máquina es de 32 bits o 64 bits. Escoge el tipo de versión de R según esto.
- Instalación de Stata: la Escuela de Ciencias Sociales proveerá un servidor donde podrás conectarte desde tu casa a la versión online de Stata. Más noticias: TBA.

Integridad Académica

- El plagio y la copia serán sancionadas con un 1. En caso de duda pregunta a tu profesor/ayudante. Procura citar todo lo que no sea de tu propiedad intelectual.
 - No se aceptan trabajos atrasados. Si tienes problemas de conectividad, planifica tus envíos con anticipación. Sólo se revisará lo que esté subido a uCampus (aunque esté incompleto). Si no hay nada, tendrás un 1.
 - Ni el ayudante ni el profesor están obligados a responder preguntas (a) después de las 5 pm durante días de semana, (b) durante fines de semana, (c) festivos.
- 📌 No existirán excepciones. Planifica tu trabajo responsablemente.

Política sobre Trabajo Cooperativo

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 trabajo final y la presentación final (más sobre esto abajo), todos los trabajos (y sus evaluaciones) serán individuales.

Evaluaciones

1. **Participación (cátedra, foro uCampus y ayudantía): 5%.**
Es fundamental que participes en clases, envíes preguntas por escrito y/o te juntes con el profesor/ayudante vía Zoom.
 - 📌 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.
2. **Control de lectura: 15%.**
Como verás en el programa, la primera sección es filosófica y conceptual. En el control de lectura se medirá cuán bien pudiste comprender los conceptos. Si tienes dudas, no dudes en contactar al ayudante/profesor.
3. **Problem Sets: 15% cada uno, 50% en total.**
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 trabajo final obligatorio/no-eximible (15%) y una presentación final (15%, vía Zoom): 35% en total.**

En este curso, la actividad final es un trabajo final (15%) que tiene formato de trabajo grupal. Usando una base de datos que nosotros te daremos, 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.** La formación del grupo es endógena.

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 (cátedra y ayudantía) | 5% | 5% |
| Problem Set #1 | 10% | 15% |
| Problem Set #2 | 10% | 25% |
| Problem Set #3 | 15% | 40% |
| Problem Set #4 | 15% | 55% |
| Control de lectura | 15% | 70% |
| Trabajo final grupal | 15% | 85% |
| Presentación grupal | 15% | 100% |

Ayudantía

Las ayudantías se harán por Zoom. Y se harán a pedido de los ayudantes. Pero en general, espera tener al menos dos ayudantías al mes.

Textos Mínimos

- Michael Ward and John Ahlquist (2001). [Maximum Likelihood for Social Science: Strategies for Analysis](#).
- Gary King (1998). [Unifying Political Methodology: The Likelihood Theory of Statistical Inference](#).
- Scott Long (1998). [Regression Models for Categorical and Limited Dependent Variables](#).
- Guido Imbens and Donald Rubin (2015). [Causal Inference for Statistics, Social, and Biomedical Sciences](#).
- Joshua Angrist and Jorn-Steffen Pischke (2009). [Mostly Harmless Econometrics: An Empiricist's Companion](#).

Textos Recomendados

- Scott Long and Jeremmy Freese (2001). *Regression Models for Categorical Dependent Variables using Stata*.
- Paul Rosenbaum (2010). *Design of Observational Studies*.

☞ También se considerarán algunos *papers*. Estos estarán señalados en las fechas indicadas y en la sección de Bibliografía.

Calendario

i. Introducción

1. Introducciones y Motivación: Por qué necesitamos MLE e Inferencia Causal?

- Introducciones: programa de curso, requerimientos y expectativas. Motivación.
- *Qué es R?* Instalación de R y RStudio.
- *Qué es Stata?*
- *Por qué necesitamos MLE e Inferencia Causal?*
- **No hay lecturas.**

ii. Inferencia Causal

2. Inferencia Causal: El Problema Fundamental en Inferencia Causal, el Supuesto de la “Ignorabilidad” y el “Potential Outcomes Framework”.

- **Lecturas:**
 - ◊ Imbens and Rubin (2015): Ch. 1.

3. Variables instrumentales y *two-stage least squares*.

- **Lecturas:**
 - ◊ Angrist and Pischke (2009): 4.1—4.2.

4. Regression discontinuity designs: *Sharp Designs*.

- **Lecturas:**
 - ◊ Angrist and Pischke (2009): 6—6.1.
 - ◊ David Lee and Thomas Lemieux (2010). *Regression Discontinuity Designs in Economics*. Journal of Economic Literature, 48(2): 281—355. Sect.: 1—3.3.

5. Regression discontinuity designs: *Fuzzy Designs*.

- **Lecturas:**
 - ◊ Angrist and Pischke (2009): 6.2.
 - ◊ Jinyong Hahn, Petra Todd and Wilbert Klaauw (2001). *Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design*. Econometrica, 69(1): 201—209.
 - ◊ David Lee and Thomas Lemieux (2010). *Regression Discontinuity Designs in Economics*. Journal of Economic Literature, 48(2): 281—355. Sect.: 3.4—7.1.
 - ◊ Guido Imbens and Karthik Kalyanaraman (2012). *Optimal Bandwidth Choice for the Regression Discontinuity Estimator*. The Review of Economic Studies, 79(3): 933—959.
 - ◊ Erik Meyersson (2014). *Islamic Rule and the Empowerment of the Poor and Pious*. Econometrica, 82(1): 229—269.

6. Incorporando el elemento *tiempo*: *fixed effects* y *differences-in-differences*.

◦ Lecturas:

- ◊ Kosuke Imai and In Song Kim (2019). *When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?* American Journal of Political Science, 62(2): 467—490.
- ◊ David Card and Alan Krueger (1994). *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*. The American Economic Review, 84(4): 772—793.
- ◊ Angrist and Pischke (2009): Ch. 5.

- Entrega temario del *Problem set* #1. Una semana de plazo. Prepárate para responder preguntas tipo ensayo breve también.

iii. Introducción a MLE: Probabilidad y Likelihood

7. Probabilidad y Likelihood.

- Diferencias filosóficas y matemáticas.
- Lecturas:
 - ◊ Ward and Ahlquist (2018): Ch. 1.
 - ◊ King (1998): Ch. 2.

8. Probabilidad e Incertidumbre.

- El enfoque de la probabilidad. Distribuciones.
- Lecturas:
 - ◊ King (1998): Ch. 3.

9. Likelihood e Inferencia.

- El enfoque del *likelihood*. OLS vía MLE?
- Lecturas:
 - ◊ King (1998): Ch. 4.1—4.3.
 - ◊ Ward and Ahlquist (2018): Ch. 4.

10. Propiedades Estadísticas y Numéricas del Likelihood.

- Finite sample problems. Precisión estadística: *Wald test*, *Likelihood Ratio*, *Lagrange Multiplier* (teórico).
- Lecturas:
 - ◊ Ward and Ahlquist (2018): Ch. 2.1—2.2.
 - ◊ King (1998): Ch. 4.4—4.6.
 - ◊ Long and Freese (2001): Ch. 2.6.

- Control de lectura. Entrada todo lo visto hasta este momento hasta el último control/*problem set*.

iv. Maximum Likelihood Estimation para Outcomes Binarios: Los Modelos Logit y Probit

11. Derivación.

- Derivando el *likelihood* del modelo logit y del modelo probit. Llegadas lineal y no-lineal. Función “link”.
- Lecturas:
 - ◊ Long and Freese (2001): Ch. 3.1—3.6.

- ◊ Ward and Ahlquist (2018): Ch. 3.1—3.4.1

12. Hypothesis testing.

- *Wald test, Likelihood Ratio, Lagrange Multiplier* (práctico).
- **Lecturas:**
 - ◊ Ward and Ahlquist (2018): Ch. 5—5.1.2; 5.1.4—5.2.
 - ◊ Long and Freese (2001): Ch. 4.1.—4.1.5.

13. Inferencia e Interpretación.

- Intervalos de confianza, *odds ratios*, *partial changes en y*, *predicted probabilities* (métodos gráficos y tablas).
- **Lecturas:**
 - ◊ Long and Freese (2001): Ch. 3.7—3.9
 - ◊ Ward and Ahlquist (2018): Ch. 6.

14. Diagnósticos.

- Análisis de residuos (Cook's D_i , Pearson's r_i , DFBETA y DFFIT) y *goodness of fit* (pseudo- R^2 e *information criteria* BIC & AIC).
- **Lecturas:**
 - ◊ Ward and Ahlquist (2018): Ch. 5—5.1.2; 5.1.4—5.2.
 - ◊ Long and Freese (2001): Ch. 4.2—4.4.
 - ◊ Gary King (1986). *How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science*. American Journal of Political Science, 30(3): 666—687.

📅 Entrega temario del *Problem set* #2. Una semana de plazo.

v. Maximum Likelihood Estimation (MLE) para Modelos Generalizados (GLMs)

15. Presentando los GLMs. *Outcomes* Ordenados: Ordered Logit/Probit.

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Ward and Ahlquist (2018): Ch. 7.
 - ◊ Long and Freese (2001): Ch. 5.
 - ◊ Ward and Ahlquist (2018): Ch. 8.

16. *Outcomes* Desordenados o “Nominales”: Multi-Nomial Logit/Probit.

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Long and Freese (2001): Ch. 6.
 - ◊ Ward and Ahlquist (2018): Ch. 9—9.3.2.

📅 Entrega temario del *Problem set* #3. Una semana de plazo.

17. *Outcomes* de Cuentas: Modelos Poisson y Negative-Binomial.

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Long and Freese (2001): Ch. 8—8.3.4.
 - ◊ Ward and Ahlquist (2018): Ch. 10.

vi. Extensiones

18. **Outcomes Poco Frecuentes: Rare Event Logistic y Zero-Inflated Poisson.**

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Long and Freese (2001): Ch. 8.5—8.6.
 - ◊ King and Zeng (2001).

■ Entrega temario del *Problem set* #4. Una semana de plazo.

19. **Outcomes Censurados/Truncados: Tobit Models.**

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Long and Freese (2001): Ch. 7.

20. **MLE y Causal Inference: Matching, Covariate Balance, y el Propensity Score.**

- Derivación. Estimación. Interpretación.
- **Lecturas:**
 - ◊ Ho et al. (2007).
 - ◊ Imbens and Rubin (2015): Ch. 13, 14 y 15.
 - ◊ Urdinez and Cruz (2019): Ch. 10.7.6.

■ Entrega del temario y asignación de bases de datos para el Trabajo Final. Dos semanas de plazo.

vii. Sesiones Guiadas de Trabajo y Presentaciones Finales

21. **Sesión Guiada de Trabajo #1**

- Oportunidad para trabajar en grupo bajo la supervisión del profesor/ayudante. Llevar dudas.

22. **Sesión Guiada de Trabajo #2**

- Oportunidad para trabajar en grupo bajo la supervisión del profesor/ayudante. Llevar dudas.

23. **Sesión Guiada de Trabajo #3**

- Oportunidad para trabajar en grupo bajo la supervisión del profesor/ayudante. Llevar dudas.

24. **Presentaciones Finales**

- Entrega de script en uCampus y presentación en formato conferencia “online”. Todos presentan. Grupos de entre 2 y 3 personas (máximo).

References

- Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. 392. Princeton University Press.
- Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–793.
- Hahn, Jinyong, Petra Todd, and Wilbert Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–209.
- Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.
- Imai, Kosuke, and In Song Kim. 2019. "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2): 467–490.
- Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3): 933–959.
- Imbens, Guido, and Donald Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30 (3): 666–687.
- . 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, 1–274. Ann Arbor, MI: University of Michigan Press.
- King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9 (2): 137–163.
- Lee, David, and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.
- Long, John, and Jeremmy Freese. 2001. *Regression Models for Categorical Dependent Variables using Stata*. College Station, Texas: STATA Press.
- Long, Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. 297. Sage.
- Meyersson, Erik. 2014. "Islamic Rule and the Empowerment of the Poor and Pious." *Econometrica* 82 (1): 229–269.
- Rosenbaum, Paul. 2010. *Design of Observational Studies*. Springer Series in Statistics. Springer New York.
- Urdinez, Francisco, and Andrés Cruz. 2019. *AnalizaR Datos Políticos*. Edited by Francisco Urdinez and Andrés Cruz. <https://arcruzo.github.io/libroadp/>.
- Ward, Michael, and John Ahlquist. 2018. *Maximum Likelihood for Social Science: Strategies for Analysis*.