

SOC3070 ANÁLISIS DE DATOS CATEGÓRICOS

Segundo Semestre 2025

Instructor:	Mauricio Bucca	Horario Clase:	Lunes 11:00 - 13:30
Email:	mebucca@uc.cl	Horario Consulta:	Lunes 14:30 - 15:30

Ayudantía:

- Ayudante: Roberto Cantillan (ricantillan@uc.cl)
- Horario: Martes 12:20 - 13:30

Páginas del curso:

- Canvas
- Repositorio Github: https://github.com/mebucca/cda_soc3070

Descripción: Este curso se centra en el estudio de métodos estadísticos para el análisis de datos categóricos, con especial énfasis en modelos de regresión para variables dependientes categóricas. Este tipo de modelos es particularmente importante dado que muchos fenómenos de interés para las ciencias sociales no se manifiestan de modo continuo sino binario (si/no), ordinal (alto/medio/bajo), nominal (candidatos A/B/C) o como recuento (números de contagios por covid-19). Cuando trabajamos con variables dependientes categóricas los modelos estándar de regresión lineal son, *por lo general*, inapropiados. En consecuencia, el curso cubrirá una amplia gama de métodos disponibles para modelar este tipo de datos, muchos de los cuales son miembros de la familia de *Modelos Lineales Generalizados* (GLM). La clase tiene como objetivo desarrollar los conocimientos y habilidades necesarias para que los estudiantes apliquen estas herramientas en su propia investigación empírica, tanto aplicada o académica.

Resultados esperados: Al completar este curso *satisfactoriamente* se espera que los estudiantes sean capaces de:

- Entender los fundamentos estadísticos y matemáticos de los principales modelos para el análisis de datos categóricos.
- Implementar estos métodos en el software estadístico **R** (o **Stata**).
- Interpretar sustantivamente los resultados de estos métodos.
- Manipular matemáticamente los resultados de estos métodos para producir "cantidades de interés".

Prerequisitos sugeridos:

- Un curso de introducción a la estadística de nivel pregrado.
- Un curso de introducción al análisis de regresión de nivel pregrado.
- Familiaridad con el software estadístico R (o Stata)

Software: Las clases y evaluaciones del curso involucrarán el uso de software estadístico. El contenido expositivo del curso y ayudantías se desarrollará exclusivamente en R. No obstante lo anterior, los estudiantes tienen libertad para desarrollar las evaluaciones tanto en R como Stata. El código de R que acompaña cada clase será puesto a disposición de los estudiantes.

Contenidos:

1. Fundamentos de teoría de la Probabilidad
2. Variables Aleatorias y Distribuciones Discretas
3. Estimación via Maximum Likelihood (MLE)
4. Tablas de Contingencia
5. Modelo Lineal de Probabilidad (LPM)
6. Modelos Lineales Generalizados (GLM)
7. Modelos de Regresión Logística
8. Modelos de Regresión Logística Multinomial
9. Modelos de Regresión Poisson y Quasi-Poisson
10. Modelos Log-Lineales para tablas de contingencia

Evaluaciones: Tareas cortas (24%), Trabajos (40%), Trabajo Final (30%), Participación (6%).

- Tareas cortas: 4 tareas cortas con una ponderación de 6% cada una.
- Trabajos: 2 trabajos de mediana extensión. Cada trabajo tiene una ponderación de 20%
- Trabajo final: 1 trabajo final con ponderación de 30% consistente en aplicar contenidos del curso a un problema sustantivo.
- Participación activa en clases y ayudantías

Bibliografía: El curso es auto-contenido y no considera un texto obligatorio. No obstante, se recomiendan los siguientes libros como textos de consulta.

- Agresti, Alan. (1996). Introduction to Categorical data Analysis. Second Edition. Wiley Series.
- Gelman, A. y Hill, J. (2007). Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press.
- Bilder, Christopher R., and Thomas M. Loughin (2014). Analysis of categorical data with R. CRC Press [Disponible online por medio de Biblioteca UC].

Políticas del curso:

- **Asistencia:** La asistencia regular a clases es importante, esperada, pero no obligatoria.
- **Entregas:** Las tareas y trabajos deben ser entregados a mas tardar el día indicado en el calendario del curso, a la hora señalada.
- **Email:** Escribir primariamente para coordinar reunión en el horario de consulta. Incluir SOC3070 en el asunto del correo. Respondo correos una vez al día (en la mañana), de lunes a viernes.
- **Honestidad Académica:** El desconocimiento de la política de honestidad académica de la universidad - incluyendo el uso de IA - no es una explicación razonable para su incumplimiento.
- **Uso de IA:** Este curso permite el uso ético y transparente de herramientas de inteligencia artificial generativa, únicamente como herramienta de estudio o apoyo en la escritura de código. El estudiantado deberá declarar explícitamente en los anexos de cada trabajo qué herramientas fueron utilizadas, con qué propósito y, si se solicita, compartir los comandos o prompts empleados. El uso de IA no exime a los y las estudiantes de dominar plenamente el contenido de sus trabajos. Para verificar esto, en la clase posterior a la entrega de cualquier tarea o trabajo, un estudiante serán seleccionado al azar para explicar oralmente frente a la clase el procedimiento seguido en alguna de sus respuestas. En base a esta exposición, su nota podrá ser reconsiderada. El uso no autorizado o no declarado será considerado una falta a la integridad académica.