

SOC3070 ANÁLISIS DE DATOS CATEGÓRICOS

Segundo Semestre 2020

Instructor:	Mauricio Bucca	Horario:	Lunes 10:00 – 12:50
Email:	mebucca@uc.cl	Lugar:	Zoom

Páginas del curso:

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

Horario de consulta: Viernes 9:00 – 10:00

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 (sí/no), ordinal (alto/medio/bajo), nominal (candidatos A/B/C) o como recuento (números de contagios por covid-19). Cuando trabajamos con variables dependiente 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 éste tipo de datos, muchos de los cuales son instancias especiales de los así llamados *Modelos Lineales Generalizados* (GLM). La clase tiene como objetivo desarrollar los conocimientos y habilidades necesarias para que los estudiantes apliquen éstos métodos en su propia investigación empírica.

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 éstos métodos en el software estadístico **R** (o **Stata**).
- Interpretar sustantivamente los resultados de éstos métodos.
- Manipular matemáticamente los resultados de éstos 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 se desarrollará exclusivamente en **R**, sin embargo los estudiantes tienen libertad para desarrollar las evaluaciones tanto en **R** como en **Stata**. El código de **R** que acompaña cada clase será puesto a disposición de los estudiantes.

Contenidos:

- Fundamentos de teoría de la Probabilidad
- Variables Aleatorias y Distribuciones Discretas
- Estimación via Maximum Likelihood (MLE)
- Tablas de Contingencia
- Modelo Lineal de Probabilidad
- Modelos Lineales Generalizados
- Modelos de Regresión Logística
- Modelos de Regresión Logística Ordenada
- Modelos de Regresión Logística Multinomial
- Modelos de Regresión Poisson y alternativas* (Regresión Binomial Negativa)
- Modelos Log-lineales para tablas de contingencia*

* Si el tiempo lo permite.

Evaluaciones: Tareas cortas (30%), Trabajos (40%), Trabajo final (30%).

- Tareas cortas: Número indeterminado de tareas cortas con ponderación variable. La ponderación de cada tarea se comunicará al momento de la entrega. 30%
- Trabajos: Dos trabajos de mediana extensión. Cada trabajo tiene una ponderación de 20% 40%
- Trabajo final: Trabajo consistente en replicar resultados seleccionados de un artículo publicado en una revista científica. El artículo será elegido por el profesor y los datos serán provistos a los estudiantes. Los estudiantes debe replicar los resultados, reportarlos en formato de informe y entregar un código de replicación. 30%

Políticas del curso:

- **Asistencia:** La asistencia regular a clases es importante, esperada, pero no obligatoria.
- **Email:** Escribir primariamente para agendar reunión en el horario de consulta. Incluir SOC3070 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 no es una explicación razonable para su incumplimiento.