

Análisis Avanzado de Datos Sociales usando R
Estación Lastarria – 2020
Docente: *Felipe Ruiz Bruzzone*

Descripción

El presente curso tiene como principal objetivo entregar una aproximación general a técnicas avanzadas de análisis de datos sociales utilizando el lenguaje de programación R. A lo largo de las sesiones se abordarán los diferentes elementos del trabajo de modelamiento estadístico de la realidad social (supuestos teóricos y matemáticos, pasos de análisis, su ejecución mediante RStudio, interpretación y evaluación de resultados y ejemplos de aplicación) así como el trabajo con bases de datos y diseños muestrales complejos. Todo ello considerando fenómenos de interés para diversas disciplinas de las ciencias sociales.

Durante el curso se irán compartiendo archivos de datos, sintaxis y referencias utilizadas por lo que cada estudiante se llevará un material ya construido en base al cual podrá continuar aprendizajes a la vez que implementar rápidamente el uso de técnicas de análisis avanzado de datos sociales usando R y RStudio para sus propios fines de investigación, sea en el área académica o profesional.

Responsable del taller

Felipe Stefano Ruiz Bruzzone. Sociólogo y Magíster en Ciencias Sociales por la Universidad de Chile. Profesor de diferentes asignaturas de pre y postgrado vinculadas al análisis estadístico de datos, el aprendizaje de lenguaje de programación R y el diseño de procesos de investigación en la Facultad de Ciencias Sociales de la Universidad de Chile. Sus temas de interés se articulan en torno al análisis sobre el desarrollo económico en América Latina y el área conocida como Computational Social Sciences. En el ámbito de la investigación se ha vinculado con proyectos de alto nivel académico y profesional desarrollados por prestigiosas instituciones internacionales, dependencias del Estado y del sistema de investigación científica en Chile (Milenio, Fondecyt). Ha sido becado para estudios de postgrado por el CONICYT en Chile y el CONICET en Argentina. Coautor del manual [*RStudio para Estadística Descriptiva en Ciencias Sociales*](#) publicado en su segunda edición durante 2019 y presentado en la versión 2019 de la Conferencia Internacional LatinR.

- Mail de contacto: felipe.ruiz@ug.uchile.cl

Horario y realización

Las sesiones se realizarán los Martes (7 sesiones), desde el martes 12 de Mayo de 19 a 21 horas. Esto será a través de la plataforma Hangouts Meet, herramienta de Google para celebrar reuniones remotas, permitiendo interactuar a un expositor con el resto de los participantes.

Público objetivo

Estudiantes de pregrado (cuarto año en adelante) y postgrado del área de las Ciencias Sociales, Humanidades, y Ciencias de la Empresa, Marketing y Administración. Profesionales de las mismas áreas mencionadas.

Objetivos de aprendizaje

Objetivo general:

Aproximar de forma general a las y los estudiantes a diferentes técnicas estadísticas de análisis multivariable de orden exploratorio, a la gestión avanzada de bases de datos, así como al manejo de datos provenientes de diseños muestrales complejos en el ámbito de la investigación social, utilizando el lenguaje de programación Rstudio. Se espera que al final del curso cada estudiante haya integrado los conocimientos necesarios para la posterior aplicación autónoma de tales herramientas estadísticas en sus propias instancias de investigación académica y/o profesional.

Objetivos de aprendizaje específicos:

- Comprender los fundamentos epistémicos y matemáticos que sostienen la construcción de modelos de análisis estadístico multivariantes, para su aplicación con lenguaje R e interpretación sustantiva con base en el campo de las Ciencias Sociales.
- Manejar herramientas para la reestructuración, reordenamiento y reescalamiento de bases de datos, así como aplicaciones que permitan superar los límites de procesamiento computacional de dispositivos de uso personal.
- Aplicar las propiedades de diseños muestrales complejos en el análisis de datos sociales.
- Comunicar al público general (especializado y no especializado) los resultados de modelos estadísticos multivariantes, considerando sus potencialidades y límites como instrumento de interpretación de la realidad.
- Indicar bibliografía e insumos educativos generales para el uso de R y RStudio y literatura especializada sobre las técnicas de análisis estadístico revisadas.

- **Requisitos mínimos**

- Se requiere un conocimiento general de metodología y técnicas de investigación cuantitativa. En términos de contenidos específicos se requiere contar con un conocimiento intermedio en el área de la estadística social univariada y bivariada, así como en el diseño de muestras.
- Cada estudiante deberá contar con un computador portátil operativo. Es deseable tener un manejo intermedio en instalación y configuración de softwares (descarga de archivos, instalación y configuración de programas, etc.).
- Debido a que se trata de un curso de continuidad en la parrilla de cursos de Estación Lastarria, se espera que las y los estudiantes cuenten con las competencias básicas para usar lenguaje de programación R orientado al análisis de datos sociales. Específicamente: lógica general en uso de sintaxis; lectura y manejo de bases de datos (desde formato CSV, planilla Excel, SPSS y Stata); gestión de paquetes de funciones especializadas; cálculo de estadísticos univariados y bivariados y construcción de gráficos, con un manejo fluido de la familia de paquetes *tidyverse* (particularmente *dplyr* y *ggplot2*); uso básico de RMarkdown.
- Es deseable un manejo intermedio del idioma inglés.

Contenidos

Tabla 1: Contenidos por sesión

Sesión	Contenidos	Duración
1. Software libre y análisis avanzado de datos sociales	Presentación general del curso. Modelos de análisis estadístico, perspectiva multivariada. El enfoque exploratorio versus el enfoque explicativo: clasificación general de las técnicas de análisis estadístico. Panorámica general de los supuestos matemáticos de la estadística multivariada. Tipos de fuentes de información: perspectiva convencional y formatos emergentes. Software libre y presentación general del lenguaje de programación R: herramientas básicas para el manejo de datos y análisis estadísticos usando RStudio.	2 horas cronológicas
2. Análisis de correspondencias múltiples para agrupar atributos	Antecedentes históricos, supuestos teóricos y aplicaciones de la técnica en investigación social. Supuestos estadísticos y pasos del análisis: elección de variables, tablas de contingencia, perfiles, centro de gravedad, inercia y ejes ortogonales. Interpretación de principales resultados e indicadores para su validación. Aplicación de la técnica utilizando RStudio.	
3. Análisis de conglomerados para la segmentación de datos	Antecedentes históricos, supuestos teóricos y aplicaciones de la técnica en la investigación social. Supuestos estadísticos y pasos del análisis: elección de variables y método de conglomeración. Interpretación de principales resultados e indicadores para su validación. Aplicación de la técnica utilizando RStudio.	
4. Gestión avanzada de bases de datos y diseños muestrales complejos en R	Funciones para redimensionar y combinar bases de datos. Recursos en línea para lidiar con insuficiencias de hardware y potencia de computación. Panorámica del diseño de muestras para estudios sociales: diseños muestrales simples versus diseños muestrales complejos. Conceptos clave para la definición de un diseño muestral. Introducción al manejo de datos sociales provenientes de diseños muestreos complejos en RStudio.	
5. Análisis de modelos de regresión lineal para la explicación de fenómenos	Antecedentes históricos, supuestos teóricos y aplicaciones de la técnica en la investigación social. Supuestos estadísticos y pasos del análisis: coeficientes de correlación lineal, estructura y supuestos del modelo de regresión lineal, probabilidades y cocientes de probabilidades, elección de variables, estimación de parámetros y evaluación de los modelos. Interpretación de principales resultados e indicadores para su validación. Aplicación de la técnica utilizando RStudio.	
6. Análisis factorial exploratorio para la reducción de dimensiones en el análisis	Antecedentes históricos, supuestos teóricos y aplicaciones de la técnica en la investigación social. Supuestos estadísticos y pasos del análisis: perspectiva exploratoria frente a confirmatoria, comunalidad y varianza explicada, especificación e identificación de los modelos, estimación de parámetros, rotación de factores, indicadores de ajuste y reespecificación del modelo. Interpretación de principales resultados e indicadores para su validación. Aplicación de la técnica utilizando RStudio.	
7. Trabajo colaborativo y comunicación de resultados en investigación social	Flujo de trabajo en investigación social cuantitativa desde la perspectiva de la Ciencia Abierta. Plataformas en línea para el trabajo colaborativo y el control de versiones (Open Science Framework, control de versiones mediante controladores Git/Github). Reporte de datos automatizado utilizando RMarkdown y Bookdown orientado al reporte de modelos de análisis multivariado.	

Fuente: elaboración propia.