

Propuesta de curso de posgrado a dictarse en la Facultad de Ciencias Sociales de la Universidad Nacional de Córdoba.

Ciencia de Datos en Sociedad

Aplicaciones del manejo de datos cuantitativos en R para las Ciencias Sociales

Equipo docente.

Docente responsable: Juan Pablo Carranza

Equipo docente: Juan Pablo Carranza, Pablo Sebastián Gómez, Eduardo León Bologna

Carga horaria: 40 horas

Fundamentación.

En la era digital actual, la capacidad de analizar y comprender grandes volúmenes de datos es una habilidad esencial en casi todos los campos del conocimiento, y las ciencias sociales no son la excepción. Los investigadores, profesionales y estudiantes de disciplinas como sociología, ciencia política y otras áreas afines se encuentran cada vez más con la necesidad de manejar datos complejos para extraer conclusiones significativas y respaldar decisiones informadas. El análisis de datos proporciona una base sólida para el desarrollo de teorías, la evaluación de políticas públicas y la comprensión de fenómenos sociales.

El curso "Ciencia de Datos en Sociedad" se ha diseñado con el objetivo de brindar a los participantes las herramientas y conocimientos necesarios para abordar una amplia gama de datos y análisis dentro del ámbito de las ciencias sociales. A lo largo de este curso, los participantes aprenderán desde los conceptos básicos de manejo de datos hasta técnicas avanzadas de análisis y visualización. Este enfoque integral les permitirá comprender y aplicar métodos estadísticos y técnicas de visualización de datos, fundamentales para interpretar y comunicar hallazgos de manera efectiva.

Un aspecto central del curso es el análisis de encuestas y datos relacionados con la población, utilizando la Encuesta Permanente de Hogares como estudio de caso. Los participantes aprenderán a trabajar con distribuciones univariadas y bivariadas, a calcular medidas descriptivas y de asociación, y a crear visualizaciones que faciliten la interpretación de los datos. Estas habilidades son cruciales para cualquier investigador social que desee explorar y entender las características y comportamientos de diferentes grupos poblacionales.

Además, el curso dedica un espacio importante al análisis de datos geográficos, un componente esencial en la investigación social actual, al permitir ligar problemáticas sociales al territorio. Los participantes aprenderán sobre los diferentes tipos de datos geográficos, cómo manipular bases de datos geográficas y realizar operaciones espaciales como unión espacial, intersección y análisis de vecinos más cercanos. La capacidad de integrar y analizar datos geográficos permite a los investigadores abordar preguntas complejas relacionadas con la distribución territorial de fenómenos sociales y la interacción entre el territorio y la sociedad.

El curso se ha diseñado con el objetivo de brindar a los participantes las herramientas y conocimientos necesarios para utilizar R, uno de los lenguajes de programación más potentes y versátiles para el análisis de datos. A lo largo de este curso, los participantes aprenderán a interactuar con este lenguaje desde sus conceptos básicos hasta técnicas avanzadas de análisis y visualización de datos, todo ello contextualizado en ejemplos y aplicaciones reales en el ámbito de las ciencias sociales. El lenguaje de programación R es conocido por su capacidad para manejar grandes conjuntos de datos, su amplia gama de paquetes y librerías especializadas, y su comunidad activa de usuarios y desarrolladores. Estas características hacen de este lenguaje una herramienta atractiva para aquellos que buscan iniciarse en el análisis de datos, adquiriendo los conocimientos fundamentales para continuar su aprendizaje por sí mismos, apoyados por una comunidad de usuarios sumamente colaborativa.

El curso de posgrado pone un énfasis especial en la aplicación práctica de los conocimientos adquiridos. Los participantes trabajarán con datos reales, resolverán problemas concretos y desarrollarán proyectos que reflejen situaciones y desafíos propios en las ciencias sociales. Se destina un espacio considerable al manejo de datos geográficos, dada la relación íntima que liga a las ciencias sociales con el territorio, apuntando a que los participantes adquieran las herramientas necesarias para visualizar e informar este tipo de información de manera clara y atractiva. De esta manera, no solo se busca que los participantes se familiaricen con técnicas de análisis de datos, sino que también adquieran una perspectiva crítica sobre cómo utilizar datos para obtener conclusiones valiosas y contribuir al conocimiento en sus respectivas áreas.

Este curso se desarrolla en el marco del plan estratégico Sociales 4.0, y tiene como objetivo fortalecer las capacidades de la FCS en el uso de lenguajes de programación de código abierto y de software libre para analizar datos provenientes de los propios trabajos de campo, o los que se encuentran disponibles, con formato abierto. Se prevé presentar una introducción al análisis de datos de encuestas, su visualización y el tratamiento de datos georeferenciados.

La propuesta tiene como primeros destinatarios a docentes y estudiantes avanzados de las carreras de grado de la FCS. A fin de aportar a la sustentabilidad financiera, la formación se abre a públicos externos.

Objetivos.

- Asumir posición crítica respecto de la calidad y de las condiciones de producción de la información cuantitativa sobre procesos sociales, sea proveniente de encuestas o de registros.
- Reconocer la importancia de trabajar con lenguajes de programación y software libres.
- Compartir el criterio de reproducibilidad de los resultados como buena práctica en investigación.
- Describir y analizar información estructurada con formato de matriz de datos, usando el lenguaje de programación R.
- Interpretar resultados provenientes de la aplicación de diferentes técnicas de análisis.
- Incorporar herramientas para la representación de información cuantitativa de manera visual.
- Introducirse en el manejo de distintos tipos de datos geográficos para la representación territorial de la investigación social.
- Incorporar herramientas para la representación de datos geográficos en mapas estáticos y en mapas interactivos.
- Adquirir los conocimientos necesarios para continuar de manera autogestiva la exploración de las posibilidades del ecosistema R, aplicado a los propios problemas de investigación.

Contenidos.

1. Fundamentos del lenguaje de programación R.
 -  Instalación local de R y RStudio.
 -  Creación de un proyecto en R.
 -  Script, consola, entorno, paquetes y ayudas.
 -  Escribir y ejecutar comandos.

- Clases de objetos y sus atributos.
- 2. Encuestas y análisis de datos.
 - Encuesta Permanente de Hogares:
 - Población de referencia.
 - Variables relevadas.
 - Estructura de la base individual.
 - Distribuciones univariadas y medidas descriptivas.
 - Distribuciones bivariadas y medidas de asociación.
 - Visualizaciones.
- 3. Datos sociales y territorio.
 - Tipos de datos espaciales:
 - datos geográficos en formato vectorial (puntos, líneas, polígonos).
 - datos geográficos en formato ráster.
 - Manipulación de bases de datos geográficas.
 - Operaciones con objetos geográficos:
 - Unión espacial.
 - Intersección.
 - Vecinos más cercanos.
 - Otras operaciones espaciales.
 - Visualización de datos geográficos:
 - Creación de mapas estáticos
 - Creación de mapas interactivos.

Bibliografía.

- Bologna, E. (2022). Un recorrido por los métodos cuantitativos de investigación a bordo de R. Disponible en:

<https://estadisticacienciasocialsr.rbind.io/>

- Bowers, J., Voors, M. & Ichino, N. (2022) (Traducido por Lily Medina). La teoría y la práctica de los experimentos de campo: Una introducción de los Learning Days de EGAP Módulo 12. Disponible en:
https://egap.github.io/theory_and_practice_of_field_experiments_spanish/introducci%C3%B3n-a-r-yrstudio.html
- comunidad de R de Latinoamérica (2023). *R Para Ciencia de Datos*, disponible en <https://es.r4ds.hadley.nz/>. Es traducción de la primera edición de *R for Data Science* de Hadley Wickham y Garrett Golemund (2017), disponible en: <https://r4ds.had.co.nz/>
- Mendoza Vega, J. B. (S/F). *R para principiantes*. Disponible en: <https://bookdown.org/jboscomendoza/rprincipiantes4/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Tercera edición. Disponible en: <https://ggplot2-book.org/>
- Pebesma, E. & Bivand, R. (2023). “Spatial Data Science, with Applications in R”. Chapman and Hall Ed. <https://doi.org/10.1201/9780429459016>. Disponible en: <https://r-spatial.org/book/>
- Notas del equipo docente.

Modalidad de dictado.

El curso se desarrollará íntegramente de manera remota, por medio de encuentros sincrónicos. Se prevé un dictado de ocho clases a razón de dos por semana, para un periodo total de cursado de 4 semanas. La carga horaria total de los encuentros sincrónicos consiste en 20 horas. Se prevé, además, una carga horaria de 20 horas adicionales para la realización de actividades asincrónicas.

La distribución de las clases obedece a la necesidad de contar con tiempo entre clases para la aplicación de procedimientos por parte del grupo cursante, dado el carácter práctico de los contenidos que se incluyen en el programa. La dedicación mínima prevista para estas actividades extra-áulicas es de 2.5 horas entre clases, en ese período, el equipo docente estará disponible para evacuar dudas en el uso de los procedimientos.

Evaluación.

La evaluación consistirá en un trabajo individual en que se deberá seleccionar un conjunto de datos a partir del cuál se solicitará identificar las características de su producción (fuente, organismo que lo financia, tipo de muestreo, entre otras) y aplicar algunos de los procedimientos de análisis del curso para responder a una pregunta contextualizada. El curso se dará por aprobado obteniendo en esta instancia una nota mínima de 7 puntos.

Fechas previstas para cada encuentro sincrónico.

- 1er. encuentro: lunes 02/09/2024, 17 a 20 hs.
- 2do. encuentro: viernes 06/09/2024, 17 a 20 hs.
- 3er. encuentro: lunes 09/09/2024, 17 a 20 hs.
- 4to. encuentro: viernes 13/09/2024, 17 a 20 hs.
- 5to. encuentro: lunes 16/09/2024, 17 a 20 hs.
- 6to. encuentro: viernes 20/09/2024, 17 a 20 hs.
- 7mo. encuentro: lunes 23/09/2024, 17 a 20 hs.
- 8vo. encuentro: viernes 27/09/2024, 17 a 20 hs.