

Análisis de datos con software libre

Equipo docente

Docente responsable: Eduardo León Bologna

Carga horaria: 60 horas, 24 presencial remotas y 36 de trabajo autónomo

Fundamentación

En la era digital, la capacidad de analizar y comprender grandes volúmenes de datos es una habilidad esencial en casi todos los campos del conocimiento. Investigadores, profesionales y estudiantes de Psicología 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 individual, la decisión sobre intervenciones y la comprensión de procesos individuales y colectivos.

La actual disponibilidad de herramientas basadas en la inteligencia artificial (IA) ha tornado más accesible procedimientos que antes estaban reservados a expertos, sin embargo el uso competente de esas herramientas demanda un base de conocimiento sobre la cual formular consultas que sean pertinentes y que conduzcan a devoluciones como las que se espera. La “ingeniería de prompts” está ganando espacio entre las habilidades requeridas para hacer uso de las nuevas facilidades.

El trayecto formativo que se propone se ha diseñado con el objetivo de brindar a quienes lo cursen las bases necesarias para abordar conjuntos de datos en diferente formato y aplicar procedimientos de análisis acordes a objetivos planteados. Se espera que quienes culminen este trayecto puedan realizar análisis descriptivos e inferenciales clásicos y que conozcan la existencia de técnicas específicas para problemas de gran complejidad, es en ese tipo de situaciones en que los recursos provistos por la IA son de ayuda.

Se trabajará desde los conceptos básicos de manejo de datos, hasta técnicas avanzadas de análisis y visualización, con el foco puesto en los fundamentos lógicos de los procedimientos, implementados en el lenguaje R. No se trata de aprender programación, sino de conocer qué herramientas están disponibles y alcanzar las competencias para hacer un uso responsable y eficaz de las facilidades que proveen los sistemas de IA para la aplicación de técnicas complejas.

Cada contenido se presenta en sus fundamentos conceptuales y se ilustra aplicándolo a conjuntos de datos abiertos, usando el lenguaje R. Se prevén usar las bases públicas de diferentes encuestas nacionales: sobre Consumos y Prácticas de Cuidado (ENCoPraC, 2022), de Factores de Riesgo (ENFR, 2005, 2009, 2013, 2018), sobre Salud Sexual y Reproductiva 2013 (ENSSyR), sobre Prevalencias de Consumo de Sustancias Psicoactivas (EnPreCoSP, 2008, 2011).

El curso no supone conocimientos previos de R, ni de otro lenguaje de programación; se inicia con la instalación local de ese programa y se recorren las etapas iniciales de su uso antes de aplicarlo al análisis. Se usarán paquetes específicos para cada tipo de procedimiento a implementar: ggplot2, summarytools, dplyr, tidyr. 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 y con la competencia necesaria para formular consultas a cualquier sistema de IA.

Sin embargo, 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, con atención a la manera de comunicar resultados, de manera visual y textual.

Objetivo General

Aportar al desarrollo de competencias en el análisis de datos, usando herramientas actualizadas

Objetivos específicos

que quienes completen el curso puedan:

Implementar procedimientos de análisis de información estructurada, 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.

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.

Adquirir los conocimientos necesarios para continuar de manera autogestiva la exploración de las posibilidades del ecosistema R, aplicado a los propios problemas de investigación, con apoyo en la comunidad de usuarios y el recurso a la IA

Contenidos

La descripción de los datos

Distribuciones univariadas

Medidas resumen correspondientes a cada tipo de variable

Distribuciones bivariadas

Medidas de asociación según tipo de variable

Visualización de datos: distribuciones simples y conjuntas

Inferencia estadística

Parámetros de variables categóricas y cuantitativas

Estimación de parámetros: error, riesgo, tamaño de muestra

La varianza según el diseño muestral

Pruebas de hipótesis: tipos de error, curva de potencia

Modelización

ANOVA de un factor

Regresión lineal simple y múltiple

Regresión logística

Bibliografía

- Bologna, E. (2025). Un recorrido por los métodos cuantitativos de investigación a bordo de R. Recuperado el 3/10/2025 de: <https://manual-estadistica-bologna.netlify.app/>
- comunidad de R de Latinoamérica (2023). *R Para Ciencia de Datos*, Recuperado el 8/6/2025 de: <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/>

- Pastukhov, A. (2024). *Data analysis using R for Psychology and Social Science*. Recuperado el 8/6/2025 de: <https://alexander-pastukhov.github.io/data-analysis-using-r-for-psychology/>
- Ruiz-Ruano García, A. M., & López Puga, J. L. (2016). R como entorno para el análisis estadístico en evaluación psicológica. *Papeles del Psicólogo*, 37(1), 74–79. Recuperado el 8/6/2025 de: <https://dialnet.unirioja.es/servlet/articulo?codigo=5354444>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Tercera edición. Recuperado el 8/6/2025 de: <https://ggplot2-book.org/>

Metodología

El curso se desarrolla de manera remota, por medio de encuentros sincrónicos a razón de un encuentro por semana durante doce semanas. Período de dictado: marzo, abril y mayo de 2026

Estrategia pedagógica

El cursado implica interacción permanente entre estudiantes y docente, los tiempos de exposición con carácter de clase magistral, se reducen al mínimo para poner foco en los intercambios que faciliten la apropiación de procedimientos de análisis. En los encuentros sincrónicos se presentarán los fundamentos de las técnicas, su aplicación a conjuntos de datos y cada cursante reproducirá los análisis presentados, para luego repetirlos con diferentes datos. Se propondrá una actividad de aplicación, a realizarse en la semana que media entre dos encuentros, para que cada cursante tenga la oportunidad de poner en evidencia eventuales problemas que surjan del uso de las técnicas presentadas en clase y éstos puedan ser resueltos en los encuentros.

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 siete (7) puntos.

Acreditación

El curso se acredita como aprobado a quienes cumplan con los requisitos de la evaluación.