

PROGRAMAS DE CURSOS Y ASIGNATURAS

TÍTULO DEL CURSO: Empieza con R: Curso de introducción al análisis y la visualización de datos

AÑO ACADÉMICO: 2025

CARRERA: Doctorado en Biología - Plan de Estudios Ord. N° 556/86, Modif. 557/10 y 807/17

FECHA DE DICTADO: desde: 7/04/2025 - hasta: 11/04/2025

DOCENTE/S RESPONSABLE/S: Patricia Elizabeth Garcia

DOCENTE/S COLABORADORES/S:

CARGA HORARIA TOTAL: 40

FUNDAMENTACIÓN: Este curso está enfocado en la actual demanda de uso y el análisis de datos de manera rápida y eficiente. El trabajo científico actual requiere entender el análisis, la elaboración y la difusión de la información obtenida. En este sentido, el lenguaje de programación R presenta una gran versatilidad en los procedimientos estadísticos y matemáticos disponibles, pudiendo analizar y visualizar los datos con gran facilidad. Actualmente R es el lenguaje líder en el análisis de datos, entre sus principales ventajas se pueden mencionar que es un software libre de acceso abierto lo cual lo convierte en transparente permitiendo un mayor control en los procesos por parte de los usuarios y además es accesible a todos. Sin embargo, debido a que utiliza código de programación su uso puede parecer intimidante. Este curso está destinado a participantes, principalmente doctorandos, que no tengan conocimientos previos en R, con el objetivo de que empiecen a utilizar el programa y además puedan familiarizarse con sus principales paquetes y funciones. Se espera que los participantes obtengan una idea general del potencial que presenta el lenguaje R y alentar a su uso de manera regular.

PROGRAMA ANALÍTICO: Unidad 1: Introducción

Instalación de R y RStudio. Elementos básicos de programación y de RStudio, descarga de paquetes, documentación disponible y ayuda. Estructuras y tipos de datos. Introducción a paquete "base". Generación de secuencias y el uso de "scripts".

Unidad 2: Datos

La estructura y atributos de los datos. Las matrices como objetos. La organización de datos en bases de datos "data frame". Funciones básicas. Lectura, importación y exportación de datos.

Unidad 3: Uso, manipulación y resumen de bases de datos

Funciones básicas usando data frame, introducción a paquete "tidyverse". Identificación y uso de algoritmos aplicados a la estadística descriptiva básica. Resumen de los datos usando función "summary".

Unidad 4: Representación gráfica de los datos

Introducción a la visualización de resultados. El uso de gráficos como herramienta de representación y difusión de los resultados. La sintaxis general para realizar gráficos, el uso de gráficos en dos dimensiones: histogramas, gráficos de barras, de líneas, de puntos y boxplot, entre otros. Introducción a la función "plot" y al paquete "ggplot2".

Unidad 5: Introducción a las variables y sus relaciones

Introducción al concepto de variables. Tipo de variables nominales u ordinales. El uso de modelos lineales: la regresión simple. Introducción al uso de los algoritmos y funciones más frecuentemente usados e incluidos en los paquetes "stats" y "car".

Unidad 6: Diseño y análisis de experimentos

Introducción al concepto de diseños factoriales. Ejemplos de uso del análisis de la varianza, algoritmos incluidos en el paquete "rstatix". Pruebas a posteriori y visualización y difusión de los resultados obtenidos mediante distintos tipos de gráficos. Uso de funciones que presentan los paquetes "ggpubr" y "ggplot2".

Unidad 7: Exploración de múltiples variables

Introducción al análisis multivariado, el uso del análisis de componentes principales (PCA) como herramienta de exploración de datos. Introducción al paquete "FactoMiner" y "factoextra". Manejo de datos faltantes con "missMDA". Representación de los resultados obtenidos mediante biplots.

Unidad 8: Principales errores en R

Ejemplos de los errores más frecuentes en el uso de RStudio. Errores de cierre ó "closure" y errores más comunes de sintaxis. Consejos y sugerencias para la búsqueda de soluciones.

OBJETIVOS: El objetivo general de este curso es que los alumnos/participantes puedan iniciarse en el uso del programa gratuito y de código abierto R. Este programa utiliza un lenguaje de programación que al principio puede no ser amigable para los usuarios. Se proporcionará a los participantes las bases para que entiendan la estructura de R, y se familiaricen con su sintaxis. Finalmente, este curso permitirá a los usuarios poder leer, escribir, manipular y analizar datos cuantitativos y cualitativos a través de la interfase RStudio.

Objetivos específicos:

- Aprender a instalar, configurar y usar el lenguaje R a través de RStudio
- Comprender la estructura de los distintos datos utilizados, teniendo en cuenta su organización y los distintos análisis que puede aplicarse
- Importar, manipular y exportar bases de datos generados en RStudio
- Utilización de los principales algoritmos estadísticos
- Producir gráficos para la difusión de los resultados obtenidos

ACTIVIDAD PRÁCTICA: Esta propuesta se basa en la presentación e introducción teórica de los contenidos seguidos de una práctica la cual puede realizarse completamente de manera presencial. Las clases teóricas incluirán presentaciones powerpoint con la introducción del

temario de día, generalmente por la mañana y la exposición de distintos ejemplos en RStudio. Además de posibilidad en tiempo real de poder realizar consultas.

Luego por la tarde se procederá a la práctica a través de guías de ejercicios que van de menor a mayor dificultad para poder ejercitar los distintos comandos, el uso de la sintaxis y las distintas características que presenta el lenguaje de R. Los encuentros tanto teóricos como prácticos son de gran importancia ya que serán actividades interactivas. Las guías de trabajos prácticos deberán ser resueltas en RStudio, con puestas en común cada determinada cantidad de ejercicios resueltos. Posteriormente, las guías de ejercicios prácticos resueltas estarán disponible para los alumnos. Por último, la evaluación será mediante un examen final integrador, que integra la parte teórica y la parte práctica.

EVALUACIÓN Y CONDICIONES DE ACREDITACIÓN: La modalidad de evaluación consistirá en la aprobación de un Examen final Integrador que constará de preguntas teórico-prácticas sobre el uso del lenguaje R. Los participantes accederán un archivo con distintas actividades tales como: preguntas de verdadero/falso, complete la frase y ejercicios que deben resolver. Los participantes tendrán 5 horas para la resolver el examen final y deberán haber aprobado la evaluación final con una nota igual ó superior a 7 (siete).

MODALIDAD DE DICTADO: PRESENCIALIDAD FÍSICA (presencialidad convencional): se desarrolla en edificios e instalaciones institucionales.

BIBLIOGRAFÍA: Quiroga, M. V., Capeletti, J., Carusso, S., Cochero, J., Garcia P.E., Llames, M. E., Morandeira, N. (2023) Limnología con R. Universidad Nacional de La Plata. Facultad de Ciencias Naturales y Museo. ISBN 978-950-34-2316-5.

An Introduction to R: <https://cran.r-project.org/doc/manuals/R-intro.html>

James, Witten, Hastie & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 6). New York: springer.

Efron, B., & Hastie, T. (2016). Computer Age Statistical Inference (Vol. 5). Cambridge University Press.

Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. Springer Science & Business Media.

Lawson, J. (2014). Design and Analysis of Experiments with R (Vol. 115). CRC press.

Paradis, E. 2003. R para principiantes. Institut des Sciences de l'Évolution. Université Montpellier II. Traducido al español por Ahumada Jorge, University of Kayahii. National Wildlife Health Center.

Wickham, H. (2015). R packages: organize, test, document, and share your code. " O'Reilly Media, Inc

Wickham, H. (2019). Advanced r. CRC press.

Wickham, H., & Golemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc."

