

Modelos Estadísticos Avanzados en R para Ciencias Biológicas



PRESENTACIÓN DE LA MATERIA

- **Docente a cargo:** Ing. Leonardo Gualano
- **Lugar:** Sociedad Argentina de Biología (curso remoto)
- **Modalidad:** Teórico-práctica a distancia, con uso del software estadístico R
- **Requerimiento previo:** Haber cursado una materia de estadística de nivel de grado
- **Requerimiento técnico:** Cada alumno debe disponer de una laptop. Se deberán instalar los softwares R y RStudio.
- **Carga horaria:** 25 horas
- **Días y horarios de cursada:** sábados y miércoles desde el 1/11/25 al 19/11/25 de 9:00 a 13:00 hs. Además, habrá una clase de una hora de consulta extra en horario a convenir.

OBJETIVOS DIDÁCTICOS

- Profundizar en el manejo avanzado de datos en R, incorporando herramientas eficientes para manipulación, transformación y visualización de datos en investigaciones biológicas.
- Introducir a los estudiantes en el uso de modelos estadísticos avanzados, incluyendo modelos lineales mixtos y modelado de varianzas, para el análisis de datos con estructuras complejas.
- Familiarizar a los estudiantes con herramientas para simulación de sistemas biológicos, permitiendo la generación y análisis de datos en contextos experimentales y teóricos.
- Capacitar en el uso de librerías especializadas como nlme, lme4, dplyr, tidyr y data.table, optimizando el flujo de trabajo en análisis de datos.

PROGRAMA

- **Clase 1: Manejo avanzado de datos.** Repaso rápido de estructuras de datos en R. Manipulación eficiente de datos con dplyr, tidyr y data.table. Programación básica en R. Secuencias de control (bucles, condicionales, funciones avanzadas).
- **Clase 2: Modelos lineales mixtos.** Repaso de regresión lineal. Introducción a modelos mixtos: estructura de efectos fijos y aleatorios. Implementación con nlme y comparación con lme4.

- **Clase 3:** Modelado de varianzas y simulación. Modelado de varianzas (varIdent(), varPower(), varExp()). Introducción a la simulación en R. Aplicaciones en biología y sistemas ecológicos.
- **Clase 4:** Análisis multivariado. Análisis de Componentes Principales (PCA). Clustering y clasificación. Aplicaciones en datos reales.
- **Clase 5:** Otras aplicaciones avanzadas. Estrategias para selección de modelos y validación cruzada. (Opcional) Introducción a series temporales: datos longitudinales en modelos mixtos.
- **Clase 6:** Workshop con trabajos de los estudiantes y discusión. Cierre de la materia.

EVALUACIÓN Y RÉGIMEN DE APROBACIÓN

- Para acreditar la materia, el alumno deberá aprobar un examen, con calificación numérica.

BIBLIOGRAFÍA

- Denis DJ (2020). Univariate, Bivariate and Multivariate Statistics Using R. Wiley.
- Fernández Casal R, Cao R, Costa J (2023). Técnicas de Simulación y Remuestreo.
- Jones O, Maillardet R, Robinson A (2014). Introduction to Scientific Programming and Simulation Using R. CRC Press.
- Seefeld K and Linder E (2007). Statistics Using R with Biological Examples. https://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf.
- Tattar PN, Ramaiah S, Manjunath BG (2016) A Course in Statistics with R. Wiley.
- Villagarcía T. Series Temporales. (material brindado por el docente del curso).
- Zuur AF, Ieno EN, Smith GM (2007). Analysing Ecological Data. Springer.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009). Mixed Effects Models and Extensions in Ecology with R. Springer.