

## Introducción a R con *tidyverse*, Modelos Lineales Generalizados (GLM) y datos desbalanceados

### Curso presencial

**Docentes:** Stefano Cabras y Juan Miguel Marín (Departamento de Estadística UC3M)

**Duración:** 24 horas

**Sesiones:** 15, 16, 17, 18, 22, 23, 24\* y 25 enero 2023 | 10:00-13:00 h.

\*salvo sesión día 24: 16:00-19:00 h.

**Aula:** 18.1.A11 ([Campus de Getafe](#))

**Créditos:** 2 créditos de formación transversal de doctorado

*Se requiere asistencia mínima a 6 sesiones completas para obtener los créditos.*

### Objetivos

Este curso se propone hacer una introducción completa a la programación en R centrada en el análisis de datos utilizando el paquete *tidyverse*, seguido por una exploración de Modelos Lineales Generalizados (GLM) y técnicas para manejar problemas de datos desbalanceados. Los estudiantes desarrollarán habilidades prácticas en la limpieza, manipulación, y análisis de datos, al tiempo que se familiarizarán con conceptos estadísticos clave.

### Contenidos

1. Introducción a R: Instalación, entorno de trabajo y conceptos básicos.
2. Domina la estructura de tus datos: Introducción al paquete *dplyr* en R.
3. Transforma y da forma a tus datos: Descubriendo el universo *tidyverse* en R.
4. Introducción a los Modelos Lineales Generalizados (GLM).
5. Tratamiento de datos desbalanceados: técnicas de sobremuestreo y submuestreo.
6. Casos de estudio y aplicaciones prácticas.

## Requisitos previos

No se requieren conocimientos previos de programación, aunque se recomienda tener una comprensión de conceptos básicos de Estadística.

**Los estudiantes deberán traer a clase sus propios ordenadores portátiles con R y RStudio instalado.**

## Metodología

El curso será impartido de forma presencial mediante una combinación de lecciones teóricas y sesiones prácticas. Los estudiantes tendrán acceso a un aula virtual con recursos y ejercicios para practicar entre las clases.

## Evaluación

El 75% de la evaluación se basará en la asistencia a las clases.

El 25% restante se evaluará a través de un proyecto final en el que los estudiantes aplicarán lo aprendido a un conjunto de datos reales.

## Bibliografía

- Wickham, H., & Golemund, G. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.

## Profesorado

Stefano Cabras y Juan Miguel Marín son profesores del departamento de Estadística de la UC3M con amplia experiencia en programación en R y análisis de datos.