

# Métodos Estadísticos de Predicción

---

Programas de  
Ordenador

Grado en Matemáticas

## AUTORES

- Víctor Aceña Gil
- Isaac Martín de Diego

2025-2026

Software utilizado en la asignatura.

# Índice programas de ordenador

|  |          |
|--|----------|
| <b>Software utilizado en la asignatura</b>                       | <b>3</b> |
| Laboratorio 0: Introducción al Entorno de Trabajo . . . . .      | 3        |
| Laboratorio 1: Regresión Lineal Simple . . . . .                 | 3        |
| Laboratorio 2: Regresión Lineal Múltiple . . . . .               | 4        |
| Laboratorio 3: Ingeniería de Características . . . . .           | 4        |
| Laboratorio 4: Selección de Variables y Regularización . . . . . | 4        |
| Laboratorio 5: Modelos Lineales Generalizados (GLM) . . . . .    | 4        |

# Software utilizado en la asignatura

Los materiales prácticos de esta asignatura consisten en una serie de laboratorios desarrollados en R y presentados en formato de cuaderno interactivo (.qmd de Quarto). Estos laboratorios son fundamentales para aplicar los conceptos teóricos vistos en clase, permitiendo al estudiante no solo ejecutar el código, sino también entender el razonamiento detrás de cada paso del análisis.

Para garantizar su preservación y acceso abierto, el código fuente completo de todos los laboratorios ha sido depositado como un único conjunto en el archivo de Software Heritage. Puedes consultarlo en:

SWHID: `swh:1:dir:e347e182ad5c23fdf7f3cacff0ec657dc3231478`

A continuación, se detalla el contenido de cada uno de los laboratorios incluidos en este depósito.

## Laboratorio 0: Introducción al Entorno de Trabajo

- **Descripción:** Esta sesión práctica inicial está diseñada para asegurar que todos los estudiantes tengan el entorno de R y RStudio correctamente configurado. Se realiza una introducción a las librerías fundamentales que se utilizarán a lo largo del curso, como `ggplot2` para visualización y `dplyr` para manipulación de datos, sentando las bases para los análisis futuros.
- **Fichero Fuente:** `lab0_introduccion.qmd`

## Laboratorio 1: Regresión Lineal Simple

- **Descripción:** Este laboratorio guía al estudiante a través del proceso completo de modelado con una única variable predictora. Se cubre el ajuste del modelo con la función `lm()`, la interpretación detallada de sus coeficientes y, de forma crucial, el diagnóstico completo de los supuestos del modelo mediante el análisis gráfico y estadístico de los residuos.
- **Fichero Fuente:** `lab1_regresion_simple.qmd`

## Laboratorio 2: Regresión Lineal Múltiple

- **Descripción:** Se extiende el análisis al caso multivariante. El foco principal de esta práctica es la correcta interpretación de los coeficientes bajo el principio *ceteris paribus* y el diagnóstico del problema de la multicolinealidad. Los estudiantes aprenderán a calcular e interpretar el Factor de Inflación de la Varianza (VIF) para evaluar la estabilidad del modelo.
- **Fichero Fuente:** lab2\_regresion\_multiple.qmd

## Laboratorio 3: Ingeniería de Características

- **Descripción:** Laboratorio práctico centrado en las técnicas para mejorar y flexibilizar los modelos lineales. Se aplicarán transformaciones de variables para corregir problemas de no-linealidad, se explorarán diferentes estrategias para la codificación de variables categóricas y se aprenderá a incluir e interpretar términos de interacción.
- **Fichero Fuente:** lab3\_ingenieria\_caracteristicas.qmd

## Laboratorio 4: Selección de Variables y Regularización

- **Descripción:** Esta sesión aborda el problema de seleccionar el subconjunto óptimo de predictores. Se exploran métodos clásicos de selección y se introducen las técnicas de regularización (Lasso y Ridge) como una solución moderna para manejar la alta dimensionalidad y la multicolinealidad, utilizando la validación cruzada para optimizar los hiperparámetros.
- **Fichero Fuente:** lab4\_seleccion\_variables.qmd

## Laboratorio 5: Modelos Lineales Generalizados (GLM)

- **Descripción:** El laboratorio final extiende el marco de la regresión a variables respuesta no normales. Se trabajará con dos de los GLM más importantes: la regresión logística para resultados binarios y la regresión de Poisson para datos de conteo. La práctica incluye la interpretación de Odds Ratios e Incidence Rate Ratios y la validación de los modelos.
- **Fichero Fuente:** lab5\_modelos\_generalizados.qmd