

**49208 - MÉTODOS ESTADÍSTICOS
MULTIVARIANTES**

CENTRO: 180 - Escuela de Ingeniería Informática

TITULACIÓN: 4046 - Grado en Ingeniería Física y Matemática

ASIGNATURA: 49208 - MÉTODOS ESTADÍSTICOS MULTIVARIANTES

CÓDIGO UNESCO: 1209.09 **TIPO:** Optativa **CURSO:** 4 **SEMESTRE:** 1º semestre

CRÉDITOS ECTS: 3 **Especificar créditos de cada lengua:** **ESPAÑOL:** 3 **INGLÉS:**

Enlace a la MEMORIA DE VERIFICACIÓN

<https://www2.ulpgc.es/plan-estudio/4046/40/verificacion/4>

REQUISITOS PREVIOS RESPECTO A ASIGNATURAS DE LA TITULACIÓN

Haber cursado las asignaturas de Álgebra Lineal, Probabilidad y Estadística.

CONTENIDOS TEÓRICOS, PRÁCTICOS Y DE LABORATORIO

DESCRIPCIÓN GENERAL

La asignatura introduce los principales métodos estadísticos y computacionales para el análisis de datos multivariantes, con especial énfasis en la modelización estadística, la inferencia, la predicción y las técnicas modernas de aprendizaje estadístico. Asimismo, se abordarán aplicaciones prácticas orientadas al análisis e interpretación de datos en distintos contextos científicos y profesionales.

PROGRAMA

Lección 1. Modelo lineal gaussiano (6 horas)

- Modelo normal multivariante.
- Estimación por máxima verosimilitud.
- Relación con el método de mínimos cuadrados ordinarios.
- Teorema de Gauss-Markov.
- Inferencia estadística en modelos lineales.
- Diagnóstico y validación de modelos.
- Criterios de selección de modelos.

Lección 2. Modelos lineales generalizados (7–8 horas)

- Familia exponencial.
- Funciones de enlace.
- Regresión logística: existencia del estimador de máxima verosimilitud e interpretación de los coeficientes.
- Regresión de Poisson y problema de la sobredispersión.
- Regresión binomial negativa.
- Aplicaciones en epidemiología y ciencias de la salud.

Lección 3. Métodos de regularización y validación (5–6 horas)

- Compromiso sesgo-varianza.
- Regresión Ridge.
- Regresión LASSO.
- Elastic Net.
- Validación cruzada.
- Métodos bootstrap para la estimación de la incertidumbre.

Lección 4. Reducción dimensional (5–6 horas)

- Análisis de componentes principales (PCA).
- Interpretación espectral.
- Análisis factorial.
- Técnicas de visualización de datos multivariantes.

Lección 5. Métodos de discriminación y clasificación (5–6 horas)

- Funciones discriminantes lineales y cuadráticas.
- Árboles de clasificación.
- Métodos de bagging y boosting.
- Redes neuronales.
- Aplicaciones al diagnóstico clínico y análisis de curvas ROC.

PROGRAMA DE PRÁCTICAS

El programa teórico anterior se complementará con sesiones prácticas desarrolladas en el entorno estadístico R, utilizando RStudio como interfaz de trabajo. En estas sesiones, el alumnado aprenderá a emplear las funciones y paquetes de R necesarios para aplicar los métodos estudiados al análisis de datos reales:

- Paquetes transversales para la gestión de datos, gráficos y estadística básica: tidyverse, ggplot, stats
- Modelo lineal gaussiano: paquete lm
- Modelos lineales generalizados: paquete glm
- Métodos de regularización: paquete glmnet.
- Reducción dimensional: paquete FactoMineR
- Métodos de discriminación y clasificación: paquetes MASS, rpart, xgboost, nnet, pROC

Las horas dedicadas a las prácticas (4) están incluidas dentro de la temporalización de las lecciones que componen el programa de la asignatura.

EVALUACIÓN:

Criterios y sistemas de evaluación

Criterios: la asignatura se evaluará a través de las siguientes Fuentes de Evaluación:

- FE1. Conocimiento de los conceptos teóricos de la asignatura.
- FE2. Capacidad de resolución de problemas relacionados con los contenidos teóricos.
- FE3. Manejo del paquete estadístico R para el análisis de datos multivariantes.
- FE4. Tener capacidad para la comprensión de publicaciones científicas en las que se utilizan los métodos estadísticos estudiados en este curso.

Sistema: la asignatura se evaluará utilizando las siguientes herramientas:

1. Examen de Teoría y Problemas: los criterios FE1 y FE2 se evaluarán a través de un examen que constará de dos partes:

- (PT) Parte teórica: desarrollo de una lección incluida en el programa.
- (RP) Resolución de problemas: planteamiento y resolución de varios ejercicios prácticos.

2. Trabajos y Prácticas: Durante las sesiones prácticas se propondrá la realización de trabajos o tareas (no entregables), orientados a profundizar en el conocimiento del lenguaje R en sus aplicaciones al análisis de datos, de acuerdo con los criterios FE3 y FE4. Cada estudiante elaborará un cuaderno de prácticas en el que recogerá las tareas realizadas tanto en las sesiones de aula como en los trabajos propuestos. Esta actividad será evaluada en una sesión práctica en la que cada estudiante podrá utilizar su propio cuaderno para resolver con R un ejercicio de análisis de datos (AD) específico, planteado a partir de una base de datos en formato Excel que se le proporcionará al efecto.

Criterios de calificación

El criterio de calificación es uniforme en todas las convocatorias y se fundamenta en la Calificación Global. Esta calificación se obtiene como la media ponderada de las puntuaciones obtenidas en las dos partes del examen (PT y RP) y en el ejercicio práctico de análisis de datos (AD), calculada como:

$$\text{CALIFICACIÓN GLOBAL} = 0.3 \cdot \text{PT} + 0.4 \cdot \text{RP} + 0.3 \cdot \text{AD}$$

La puntuación de cada una de las partes varía entre 0 y 10 y valorará el grado en que el estudiante haya alcanzado los resultados de aprendizaje esperados, consistentes en ser capaz de:

- Formular y estimar modelos lineales y modelos lineales generalizados.
- Aplicar técnicas de regularización y selección de variables.
- Emplear métodos de reducción dimensional para el análisis de datos complejos.
- Construir y evaluar clasificadores supervisados.
- Interpretar resultados estadísticos en contextos aplicados.
- Implementar técnicas multivariantes mediante software estadístico especializado.

La calificación global será la que conste en el acta académica. Esto significa que cuando dicha calificación sea igual o superior a CINCO puntos, se considerará aprobada la asignatura.

PLANIFICACIÓN SEMANAL

Semanas 1-3: Lección 1.
 Semanas 4-7: Lección 2
 Semanas 8-10: Lección 3
 Semanas 11-12: Lección 4
 Semanas 13-15: Lección 5

PROFESORADO

Dr./Dra. Ángelo Santana Del Pino

(COORDINADOR)

Departamento: 275 - MATEMÁTICAS

Ámbito: 265 - Estadística E Investigación Operativa

Área: 265 - Estadística E Investigación Operativa

Despacho: MATEMÁTICAS

Teléfono: 928458812 **Correo Electrónico:** angelo.santana@ulpgc.es

CV: [Información curricular del profesor](#)

BIBLIOGRAFÍA

[1 Básico] Applied multivariate statistical analysis /

Richard A. Johnson, Dean W. Wichern.

Prentice Hall,, Englewood Cliffs (New Jersey) : (1988) - (2nd ed.)

0130411388

[2 Básico] The theory of linear models and multivariate analysis

Steven F. Arnold

Wiley - (1981)

0-471-05065-2

[3 Básico] The Elements of Statistical Learning: data Mining, inference, and prediction /

Trevor Hastie, Robert Tibshirani, Jerome Friedman.

Springer,, New York ... [etc.] : (2009) - (2nd ed., 11th printing.)

978-0-387-84857-0

[4 Recomendado] An Introduction to Statistical Learning: with Applications in R /

Gareth James ... [et al.].

Springer,, New York ... [etc.] : (2013)

978-1-4614-7137-0

[5 Recomendado] Using R with multivariate statistics /

Randall E. Schumacker.

Sage,, Los Angeles ... [etc.] : (2016)

978-1-4833-7796-4