

**CENTRO:** 180 - Escuela de Ingeniería Informática

**TITULACIÓN:** 4046 - Grado en Ingeniería Física y Matemática

**ASIGNATURA:** 49189 - ESTADÍSTICA

**CÓDIGO UNESCO:** 1209      **TIPO:** Obligatoria      **CURSO:** 3      **SEMESTRE:** 1º semestre

**CRÉDITOS ECTS:** 6      **Especificar créditos de cada lengua:**      **ESPAÑOL:** 6      **INGLÉS:**

## Enlace a la MEMORIA DE VERIFICACIÓN

<https://www2.ulpgc.es/plan-estudio/4046/40/verificacion/1>

## REQUISITOS PREVIOS RESPECTO A ASIGNATURAS DE LA TITULACIÓN

Para cursar esta asignatura se precisan los siguientes conocimientos que previamente se han desarrollado en las siguientes asignaturas:

Algebra Lineal  
Fundamentos de Matemáticas I  
Fundamentos de Matemáticas II  
Análisis Matemático I  
Probabilidad.

## CONTENIDOS TEÓRICOS, PRÁCTICOS Y DE LABORATORIO

Bloque 0: Exploración y análisis de datos con R (8 horas)

-----

Este bloque se impartirá en las clases prácticas. R como herramienta interactiva para la exploración de datos: tablas, estadísticos descriptivos, gráficos. R como herramienta computacional para programación de procedimientos estadísticos.

Bloque 1: Fundamentos de Inferencia Estadística (30 horas)

-----

1. Introducción y distribuciones muestrales (6 horas)

Objetivos de la inferencia estadística. Población y muestra. Revisión de conceptos clave de probabilidad (leyes de los grandes números, teorema central del límite). Estadísticos y su distribución en el muestreo. Ejemplos: distribuciones de la media y varianza muestrales.  $\chi^2$ , t-Student, F-Fisher.

2. Estimación puntual (6 horas)

Función de verosimilitud. Concepto de estimador. Suficiencia de un estimador. Criterios de optimalidad: insesgadez, consistencia, eficiencia. Teorema de Rao-Blackwell. Información de Fisher y cota de Frechet-Cramer-Rao. Métodos de estimación: método de los momentos y método de Máxima Verosimilitud (ML). Propiedades asintóticas de los estimadores ML.

3. Bootstrap (5 horas)

Desarrollos de Edgeworth. Aproximación de la distribución muestral de un estimador mediante el remuestreo de datos (distribución bootstrap). Distribución bootstrap vs distribución aproximada por el Teorema del Límite Central. Bootstrap no paramétrico: muestreo con reemplazo de la distribución empírica. Bootstrap paramétrico: simulación de la distribución del estadístico bajo un modelo paramétrico ajustado a los datos. Estimación bootstrap del sesgo y varianza de un estimador.

#### 4. Contrastes de hipótesis (6 horas)

Contrastes de significación. Hipótesis nula y alternativa. Errores tipo I y tipo II. Nivel de significación. Test estadístico. Distribución del test bajo la hipótesis nula. Región crítica. Valor p. Hipótesis simples y compuestas. Función de potencia. Lema de Neyman-Pearson y contrastes uniformemente más potentes. Test de razón de verosimilitudes. Ejemplos de contrastes paramétricos clásicos. Contrastes bootstrap paramétricos.

#### 5. Intervalos y regiones de confianza (4 horas)

Concepto de intervalo de confianza. Nivel de confianza. Construcción de intervalos mediante el método de los pivotaes. Construcción de intervalos mediante la inversión de tests de significación. Ejemplos de intervalos para parámetros de distribuciones usuales. Regiones de confianza para parámetros vectoriales. Construcción de intervalos de confianza bootstrap. Métodos para la corrección del sesgo.

#### 6. Contrastes No Paramétricos (3 horas)

Tests de rangos (ejemplos: Wilcoxon, Mann-Whitney). Tests de permutaciones. Tests de bondad de ajuste. Contrastes bootstrap no paramétricos.

### Bloque 2: Modelos de análisis de datos (18 horas)

-----

#### 7. Modelos Lineales (8 horas)

Definición de modelo lineal  $Y=Xb+e$ . Supuestos básicos. Ejemplos. Matriz de diseño. Modelos de rango completo y de rango no completo. Estimación del modelo lineal: Métodos de máxima verosimilitud (ML) y mínimos cuadrados (MC). Propiedades de los estimadores MMC: insesgadez, varianza, teorema de Gauss-Markov. Inferencia en el modelo lineal: contrastes sobre los coeficientes y sobre submodelos. Intervalos de confianza para coeficientes y predicciones. Comparación de modelos. Estimación bootstrap de la varianza de los coeficientes del modelo lineal. Particularizaciones del modelo lineal: Regresión, ANOVA y ANCOVA. Representación unificada mediante la matriz de diseño. Regresión lineal como caso base del modelo lineal. ANOVA como modelo lineal con variables categóricas: uso de variables indicadoras para representar niveles de factores. ANCOVA: combinación de variables categóricas y continuas, incluyendo covariables. Interacciones: términos cruzados en la matriz de diseño.

#### 8. Diagnóstico y Validación de Modelos Lineales (5 horas)

Supuestos para la inferencia: normalidad e independencia de los errores y homocedasticidad. Análisis de residuos: gráficos de residuos vs. valores ajustados, QQ-plots para normalidad. Detección de observaciones influyentes: leverage, distancia de Cook. Colinealidad: factor de inflación de varianza (VIF). Métodos de selección de variables: criterios  $R^2$  ajustado, AIC y BIC.

#### 9. Extensiones del modelo lineal: modelos lineales generalizados (GLMs) (5 horas)

Componentes de los GLMs: componente aleatorio (familia exponencial), componente sistemático ( $\eta$ ), función de enlace. Ejemplos: regresión logística, regresión de Poisson, regresión binomial negativa. Interpretación de los coeficientes: odds-ratios y riesgos relativos. Estimación por máxima verosimilitud: algoritmos iterativos (e.g., Newton-Raphson). Inferencia: pruebas de Wald,

razón de verosimilitudes, intervalos de confianza. Validación del modelo. Sobredispersión en modelos de Poisson.

### Bloque 3: Introducción a Estadística Bayesiana

-----

#### 10. Fundamentos de estadística bayesiana (4 horas)

Interpretación bayesiana vs. frecuentista de probabilidad. Teorema de Bayes: prior, verosimilitud, posterior. Familias conjugadas: beta-binomial, gamma-Poisson, normal-normal. Ejemplo práctico: Estimación bayesiana de proporción. Inferencia bayesiana básica: Intervalos de credibilidad vs. intervalos de confianza. Contrastes bayesianos. Factor de Bayes.

## EVALUACIÓN:

### Criterios y sistemas de evaluación

-----

#### Criterios 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

FE4. Tener capacidad para comprender publicaciones científicas asequibles al nivel de los alumnos.

#### Sistema de Evaluación

Teoría y Problemas (TP). Se evaluarán a través de un examen teórico práctico en el que se propondrán cuestiones teóricas y problemas correspondientes al programa de la asignatura (criterios FE1 y FE2).

Prácticas (PS): se evaluarán a través de la realización en grupos de un TRABAJO DE ANÁLISIS DE DATOS en lenguaje R (criterios FE3 y FE4).

#### Criterios de calificación

-----

El criterio de calificación es uniforme en todas las convocatorias y se fundamenta en la Calificación Global. Esta calificación se calcula como la media ponderada de las puntuaciones obtenidas en el examen teórico-práctico (85%) y en el trabajo grupal (15%). Ambas puntuaciones varían entre 0 y 10. En resumen:

$$\text{CALIFICACIÓN GLOBAL} = 0.85 \cdot \text{TP} + 0.15 \cdot \text{PS}$$

La puntuación obtenida en el trabajo grupal (PS) podrá conservarse dentro de un mismo año académico.

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

La temporalización semanal de las actividades formativas se distribuirá de modo uniforme a lo largo de las 15 semanas de duración de la asignatura. Por tanto, de acuerdo con la establecido en el apartado de metodología, la distribución semanal de actividades formativas será:

Actividad.	Horas semanales
Sesiones académicas de fundamentación	2
Sesiones académicas de interacción	1
Sesiones académicas de prácticas de aula	1
Actividad no presencial (estudio, trabajo autónomo y resolución de problemas):	6

La distribución de las lecciones por semana será, aproximadamente, la que sigue:

Semana	Lección (horas)	Prácticas (todas de 1 hora)
1	1 (4h)	
2	1-2 (2h-2h)	
3	2 (4h)	
4	3 (3h)	1
5	3-4 (2h-1h)	2
6	4 (3h)	3
7	4-5 (2h-1h)	4
8	5-6 (3h-1h)	
9	6-7 (2h-1h)	5
10	7 (4h)	
11	7 (3h)	6
12	8 (3h)	7
13	8-9 (2h-1h)	8
14	9 (4h)	
15	10 (4h)	

## PROFESORADO

**Dr./Dra. Ángel 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](mailto:angelo.santana@ulpgc.es)

**Dr./Dra. Carmen Nieves Hernández Flores**

**Departamento:** 275 - MATEMÁTICAS

**Ámbito:** 265 - Estadística E Investigación Operativa

**Área:** 265 - Estadística E Investigación Operativa

**Despacho:** MATEMÁTICAS

**Teléfono:** 928458812 **Correo Electrónico:** [carmennieves.hdezflores@ulpgc.es](mailto:carmennieves.hdezflores@ulpgc.es)

---

**[1 Básico] Bootstrap methods and their application /**

*A.C. Davison, D.V. Hinkley.*  
*Cambridge University Press,, Cambridge ; (1997)*  
*0521574714 (pb)*

---

**[2 Básico] Computer Age Statistical Inference**

*Bradley Efron, Trevor Hastie*  
*Cambridge University Press - (2016)*  
*978-1-107-14989-2*

---

**[3 Básico] Theoretical statistics /**

*D.R. Cox, D.V. Hinkley.*  
*Chapman & Hall/CRC,, Boca Raton : (2000)*  
*0412428601 (HB)*

---

**[4 Básico] Generalized linear models /**

*P. McCullagh and J.A. Nelder.*  
*Chapman & Hall,, London : (1989) - (2ª ed.)*  
*0412317605*

---

**[5 Básico] Mathematical statistics /**

*Steven F. Arnold.*  
*Prentice Hall,, Englewood Cliffs (New Jersey) : (1990)*  
*0135630991*

---

**[6 Recomendado] An introduction to the bootstrap /**

*Bradley Efron and Robert J. Tibshirani.*  
*Chapman & Hall,, New York : (1993)*  
*0412042312*

---

**[7 Recomendado] Regression modeling strategies: with applications to linear models, logistic regression and survival analysis /**

*Frank E. Harrell, Jr.*  
*Springer,, New York [etc.] : (2001)*  
*978-0-387-95232-1*

---

**[8 Recomendado] Estadística Matemática con aplicaciones /**

*John E. Freund, Irwin Miller, Marylees Miller.*  
*Pearson Education,, Mexico : (2000) - (6ª ed.)*  
*970-17-0389-8*

---

**[9 Recomendado] Applied Regression Analysis and Generalized Linear Models /**

*John Fox.*  
*Sage,, Los Angeles : (2008) - (2nd ed.)*  
*978-0-7619-3042-6*

---

**[10 Recomendado] Linear models with R /**

*Julian J. Faraway.*  
*Chapman & Hall/CRC,, Boca Raton : (2005)*  
*1584884258*

---

**[11 Recomendado] The bootstrap and edgeworth expansion /**

*Peter Hall.*  
*Springer-Verlag,, New York : (1992)*  
*0387977201*

---

**[12 Recomendado] The theory of linear models and multivariate analysis**

*Steven F. Arnold*

*Wiley - (1981)*

*0-471-05065-2*

---

**[13 Recomendado] Elements of statistical learning: data mining, inference, and prediction /**

*Trevor Hastie; colaboradores Robert Tibshirani y Jerome Friedman.*

*Springer-Verlag,, New York : (2001)*

*0387952845*

---

**[14 Recomendado] Estadística matemática con aplicaciones /**

*William Mendenhall, Dennis D. Wackerly, Richard L. Scheaffer ; traductores, Dirk Valckx Verbeeck, Arturo de la*

*Fuente Pantoja.*

*Grupo Editorial Iberoamérica,, México : (1994) - (2ª ed.)*

*9706250166*