

## **Análisis de Datos Complejos**

Código: 104399  
Créditos ECTS: 6

| Titulación                                            | Tipo | Curso | Semestre |
|-------------------------------------------------------|------|-------|----------|
| 2503740 Matemática Computacional y Analítica de Datos | OB   | 2     | 2        |

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

### **Contacto**

Nombre: Pere Puig Casado  
Correo electrónico: Pere.Puig@uab.cat

### **Uso de idiomas**

Lengua vehicular mayoritaria: catalán (cat)  
Algún grupo íntegramente en inglés: No  
Algún grupo íntegramente en catalán: Sí  
Algún grupo íntegramente en español: No

### **Otras observaciones sobre los idiomas**

Los materiales docentes que se cuelguen en el Campus Virtual serán en Inglés

### **Equipo docente externo a la UAB**

Dorota Mlynarczyk

### **Prerequisitos**

Se recomienda un buen conocimiento de la asignatura de Modelización e Inferencia así como una cierta soltura con el software R.

### **Objetivos y contextualización**

El principal objetivo es proporcionar herramientas estadísticas para el análisis de datos, dominando las técnicas más relevantes para enfrentarse con modelos complejos.

### **Competencias**

- Calcular y reproducir determinadas rutinas y procesos matemáticos con agilidad.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Formular hipótesis e imaginar estrategias para confirmarlas o refutarlas.
- Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
- Utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, visualización gráfica, optimización u otras para experimentar y resolver problemas.

## Resultados de aprendizaje

1. Analizar datos mediante técnicas de inferencia para una o dos muestras.
2. Conocer los diferentes métodos de recogida de datos.
3. Escoger el software estadístico apropiado para analizar los datos mediante técnicas de inferencia.
4. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
5. Identificar la inferencia estadística como instrumento de pronóstico y predicción.
6. Identificar las distintas fuentes de información disponibles.
7. Interpretar los resultados obtenidos y concluir respecto a la hipótesis experimental.
8. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
9. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
10. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
11. Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
12. Utilizar las propiedades de las funciones de densidad.
13. Utilizar las propiedades de las funciones de distribución.
14. Utilizar software estadístico para gestionar bases de datos.
15. Utilizar software estadístico para obtener índices de resumen de las variables del estudio.
16. Validar y gestionar la información para su tratamiento estadístico.

## Contenido

1- Modelos lineales: regresión múltiple y ANOVA.

2- Modelos lineales generalizados: regresión logística y de Poisson.

3- Métodos de remuestreo 1: tests permutacionales.

4- Métodos de remuestreo 2: bootstrap.

5- Métodos de remuestreo 3: jackknife.

Si hay tiempo también se hará una introducción al Análisis de Componentes Principales.

## Metodología

De acuerdo con los objetivos de la asignatura, el desarrollo del curso se basa en las siguientes actividades:

Clases de teoría: El alumno adquiere los conocimientos científico-técnicos propios de la asignatura asistiendo en las clases de teoría complementándolas con el estudio personal de los temas explicados. Las clases de teoría son las actividades en las cuales se exige menos interactividad al estudiante: están concebidas como un método fundamentalmente unidireccional de transmisión de conocimientos del profesor al alumno. Las

clases se harán utilizando un apoyo de diapositivas PowerPoint en Anglès que se colgarán también al Campus Virtual.

**Problemas y prácticas:** Los problemas y las prácticas son sesiones con una doble misión. Por un lado se trabajan los conocimientos científico-técnicos expuestos en las clases de teoría para completar su comprensión y profundizar en ellos desarrollando actividades diversas, desde la típica resolución de problemas hasta la discusión de casos prácticos. Por otro lado, las clases de problemas son el foro natural en el cual discutir en común el desarrollo del trabajo práctico, aportando los conocimientos necesarios para llevarlo adelante, o indicando donde y como se pueden adquirir.

El curso práctico de esta asignatura se plantea como un camino para orientar el estudiante en un trabajo de campo de estadística en cada una de sus etapas. Así realizan por su cuenta con el software R unos ejercicios prácticos dirigidos a resolver problemas reales concretos. Este planteamiento está orientado a promover un aprendizaje activo y a desarrollar el razonamiento crítico y la capacidad de análisis y síntesis.

**Nota:** se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

## Actividades

| Título              | Horas | ECTS | Resultados de aprendizaje               |
|---------------------|-------|------|-----------------------------------------|
| Tipo: Dirigidas     |       |      |                                         |
| Clases de Teoría    | 24    | 0,96 | 4, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16 |
| Clases de problemas | 20    | 0,8  | 4, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13     |
| Tipo: Supervisadas  |       |      |                                         |
| Sesiones prácticas  | 20    | 0,8  | 1, 4, 3, 11, 14, 15, 16                 |
| Tipo: Autónomas     |       |      |                                         |
| Trabajo personal    | 61    | 2,44 | 4, 3, 5, 6, 7, 9, 10, 12, 13, 14, 15    |

## Evaluación

La evaluación se realiza a lo largo de todo el curso. La evaluación continuada tiene varios objetivos fundamentales: Monitorizar el proceso de enseñanza y aprendizaje, permitiendo tanto al alumno como al profesor conocer el grado de logro de las competencias y corregir, si es posible, las desviaciones que se produzcan. Incentivar el esfuerzo continuado del alumno frente al sobreesfuerzo, frecuentemente inútil, de última hora. Verificar que el alumno ha logrado las competencias determinadas en el plan de estudios.

Para hacer esta evaluación se cuenta con los siguientes instrumentos: Los ejercicios prácticos librados por los estudiantes (30%), un examen parcial de Teoría en medio del curso (35%), otro examen parcial de Teoría al final del curso (35%). A la recuperación solo podrán ir los alumnos que tengan un mínimo de 3 en la nota final, recuperando solo la parte de Teoría.

## Actividades de evaluación

| Título     | Peso | Horas | ECTS | Resultados de aprendizaje |
|------------|------|-------|------|---------------------------|
| Ejercicios | 30   | 20    | 0,8  | 4, 3, 11, 14, 15          |

|                  |    |     |     |                                     |
|------------------|----|-----|-----|-------------------------------------|
| Examen parcial 1 | 35 | 2,5 | 0,1 | 1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 16 |
| Examen parcial 2 | 35 | 2,5 | 0,1 | 1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 16 |

## Bibliografía

- Introduction to Linear Regression Analysis. Montgomery, D. Peck, A. Vining, G., 2001.
- An R Companion to Linear Statistical Models. Christopher Hay-Jahans, 2012.
- Generalized Linear Models. McCullagh, P. and Nelder, J., 1992.
- Resampling methods: a practical guide to data Analysis. Phillip I. Good, 2006.
- The jackknife, the bootstrap and other resampling plans. Bradley Efron, 1982.
- Bootstrap methods and their application. A.C. Davison, D.V. Hinkley, 1997.

## Software

Se utilizará el software R.