

## Visualización de Datos y Modelización

Código: 43482  
Créditos ECTS: 6

| Titulación                                                                                   | Tipo | Curso | Semestre |
|----------------------------------------------------------------------------------------------|------|-------|----------|
| 4313136 Modelización para la Ciencia y la Ingeniería / Modelling for Science and Engineering | OT   | 0     | 1        |

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

### Equipo docente

Rosa Camps Camprubí

Rosario Delgado de la Torre

Juan Ramón González Ruíz

### Uso de idiomas

Lengua vehicular mayoritaria: inglés (eng)

### Prerequisitos

Un conocimiento elemental en Teoría de la Probabilidad y en Inferencia Estadística.

### Objetivos y contextualización

Curso de R. Todos los ejercicios prácticos se harán utilizando el paquete estadístico R. Este curso introductorio es básico por los posteriores desarrollos.

Visualización de grandes conjuntos de datos con R GViz, Maps y Tabplot.

Simulación de datos, Bootstrap y tests permutacionales Estas metodologías permiten una solución rápida para modelos estadísticos complejos sin un conocimiento profundo de los métodos estadísticos generales clásicos. Son herramientas indispensables entre las actuales técnicas de modelización estadística. Los estudiantes completarán un programa de entrenamiento básico, incluyendo el software adecuado, y aprenderán como enfrentarse a problemas con datos reales.

Redes Bayesianas, en opinión de muchos investigadores, constituyen una de las más significantes contribuciones de la IA en este siglo. Se trata de estructuras gráficas para representar relaciones probabilísticas entre un gran número de variables que sirven también para hacer inferencia probabilística con estas variables en un gran número de campos de aplicación. Uno de los objetivos de este curso es introducirlas, desarrollando en los estudiantes la habilidad de utilizarlas en modelización, tanto desde un punto de vista teórico como práctico, con un particular énfasis en la utilización de un software apropiado.

### Competencias

- "Aplicar el pensamiento lógico/matemático: el proceso analítico a partir de principios generales para llegar a casos particulares; y el sintético, para a partir de diversos ejemplos extraer una regla general."
- Analizar sistemas complejos de distintos campos y determinar las estructuras y parámetros básicos de su funcionamiento.
- Analizar, sintetizar, organizar y planificar proyectos de su campo de estudio.
- Aplicar la metodología de investigación, técnicas y recursos específicos para investigar en un determinado ámbito de especialización.
- Aplicar las técnicas de resolución de los modelos matemáticos y sus problemas reales de implementación.
- Concebir y diseñar soluciones eficientes, aplicando técnicas computacionales, que permitan resolver modelos matemáticos de sistemas complejos.
- Extraer de un problema complejo la dificultad principal, separada de otras cuestiones de índole menor.
- Formular, analizar y validar modelos matemáticos de problemas prácticos de distintos campos.

## Resultados de aprendizaje

1. "Aplicar el pensamiento lógico/matemático: el proceso analítico a partir de principios generales para llegar a casos particulares; y el sintético, para a partir de diversos ejemplos extraer una regla general."
2. Analizar, sintetizar, organizar y planificar proyectos de su campo de estudio.
3. Aplicar técnicas de Series Temporales para estudiar modelos asociados a problemas prácticos.
4. Aplicar técnicas de Series Temporales para predecir el comportamiento futuro de ciertos fenómenos.
5. Aplicar la metodología de investigación, técnicas y recursos específicos para investigar en un determinado ámbito de especialización.
6. Extraer de un problema complejo la dificultad principal, separada de otras cuestiones de índole menor.
7. Identificar los parámetros que determinan el funcionamiento de un sistema
8. Implementar las soluciones propuestas de forma fiable y eficiente.
9. Reconocer problemas que requieran aplicar técnicas de Series Temporales para construir modelos asociados a problemas prácticos.
10. Seleccionar la mejor descripción de un sistema en función de sus características particulares
11. Usar softwares específicos para la resolución de problemas de optimización.

## Contenido

Parte 1a: Introducción al R (6h)

Parte 2a: Visualización de grandes conjuntos de datos con R (6h)

Part 3a: Redes Bayesianas (14h)

- 1) Block 1: Elementos básicos.
- 2) Block 2: Redes causales y Inferencia en Redes Bayesianas.
- 3) Block 3: Aprendizaje de los parámetros de las redes bayesianas.

Parte 4a: Simulación de datos, Bootstrap y Tests Permutacionales (12h)

- 1) Tests Permutacionales.
- 2) Jackknife.
- 3) Bootstrap paramétrico.
- 4) Bootstrap no-paramétrico.

## Metodología

Las clases magistrales de este curso, en las que es determinante la explicación del profesor, son la base del proceso de aprendizaje. Es también muy importante la participación de los alumnos, combinada con sesiones prácticas en las que el estudiante tiene que usar el conocimiento aprendido para solucionar problemas.

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 magistrales | 38    | 1,52 | 3, 4, 7, 8, 9, 10                 |
| Ejercicios         | 16    | 0,64 | 1, 2, 5, 3, 4, 6, 7, 8, 9, 10, 11 |
| Proyectos+Trabajos | 18    | 0,72 | 1, 2, 5, 3, 4, 6, 7, 8, 9, 10, 11 |
| Tipo: Supervisadas |       |      |                                   |
| Sesiones prácticas | 20    | 0,8  | 3, 4, 7, 8, 9, 10                 |

## Evaluación

La evaluación del curso consiste en una evaluación continuada.

Hay 4 evaluaciones durante el curso, que tienen un peso de 10%, 10%, 40% y 40% respectivamente.

Cada profesor contará con su propio tipo de evaluación.

Evaluación de la parte 1: Trabajo continuado + proyecto final (análisis individual de datos reales con R).

Evaluación de la parte 2: Trabajo continuado + proyecto

Evaluación de la parte 3: Trabajo continuado + Entrega de algunos ejercicios + examen

Evaluación de la parte 4: Trabajo continuado + Entrega de algunos ejercicios.

## Actividades de evaluación

| Título             | Peso | Horas | ECTS | Resultados de aprendizaje         |
|--------------------|------|-------|------|-----------------------------------|
| Proyectos y Examen | 50   | 20    | 0,8  | 1, 2, 5, 3, 4, 6, 7, 8, 9, 10, 11 |
| Trabajo continuado | 50   | 38    | 1,52 | 1, 2, 5, 3, 4, 6, 7, 8, 9, 10, 11 |

## Bibliografía

- 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.
- "Learning Bayesian Networks" by R. E. Neapolitan, Prentice Hall Series in Artificial Intelligence, 2004.

- "Probabilistic Methods for Bioinformatics with an Introduction to Bayesian Networks" by R. E. Neapolitan, Elsevier, 2009.

## **Software**

Se utilizará el software R.