

## GUIA DOCENTE

### MÁSTER UNIVERSITARIO EN DATA ANALYTICS FOR BUSINESS

Edición 2

Curso 2024-2025

#### 1. ASIGNATURA

- **Nombre:** *Fundamentos de estadística para el análisis de datos*
- **Tipo de asignatura:** Optativa
- **Trimestre:** PRIMERO
- **Créditos:** 3 ECTS
- **Idioma de docencia:** castellano
- **Coordinador de la asignatura:** Jordi Pereira
- **Profesor/es de la asignatura:** Jordi Pereira

#### 2. PRESENTACIÓN DE LA ASIGNATURA

##### Objetivos de la asignatura

El objetivo principal del curso es la presentación de conceptos de estadística y probabilidad necesarios para el desarrollo y comprensión de herramientas de aprendizaje automático. Partiendo de las nociones básicas de ambas disciplinas, el curso cubre los conocimientos teóricos de estadística y probabilidad necesarios para la comprensión de estas nuevas herramientas. Por tanto, los objetivos específicos del curso son:

- (1) Dotar al alumnado de las herramientas y fundamentos de la estadística y probabilidad clásica.
- (2) Dotar al alumnado de conocimientos sobre las diferentes pruebas y procedimientos de estadística necesarios para el desarrollo de hipótesis y su verificación.
- (3) Capacitar al alumnado para identificar las limitaciones y errores asociados a la generalización derivada de observaciones

Nuestro compromiso con el impacto social y el bienestar planetario se traduce en contenidos formativos alineados con los Objetivos de Desarrollo Sostenible (ODS) previstos en la agenda 2030:



En la asignatura que nos ocupa, los ODS implicados son:

- ODS.4. Educación de Calidad
- ODS.8. Trabajo Decente y Crecimiento Económico

### Contenidos

La asignatura cubre los siguientes temas:

- (1) Conceptos básicos. Población, muestra, estimadores puntuales.
- (2) Preparación y preproceso de datos. Identificación de datos atípicos y valores faltantes.
- (3) Modelos de regresión y optimización de parámetros.
- (4) Intervalos de confianza.
- (5) Tesis de hipótesis.
- (6) Tipos de errores y métricas de evaluación.
- (7) Estrategias para la preparación de datos y técnicas de reducción de dimensionalidad.

### La asignatura dentro del plan de estudios

*Esta asignatura obligatoria/optativa se emmarca dentro de la materia 1. **Análisis de Datos. Data Analytics** del plan de estudios. Se realiza durante el **primer trimestre**.*

### Competencias/Resultados de aprendizaje

RA1. Mat 1.1 Seleccionará la infraestructura necesaria para hacer frente a un proyecto que involucre datos masivos.

RA3 Mat 1.2 Realizará un informe completo que incluya visualizaciones diversas sobre un conjunto de datos y que ayuden a la toma de decisiones.

RA4. Mat 1.3 Extraerá información de modo visual a partir de datos masivos.

RA5. Mat. 1.4 Identificará las diferencias fundamentales entre bases de datos relacionales y no relacionales.

RA6. Mat 1.5 Analizará un conjunto de datos mediante análisis univariante y bivariante.

RA6. Mat 1.6 Distinguirá correlaciones entre las diferentes dimensiones de un conjunto de datos.

RA7. Mat 1.7 Realizará un plan de gobernanza de datos que minimice riesgos y costes.

RA19. Mat 1.1 Utilizará python para generar visualizaciones adecuadas al tipo de datos que se estén trabajando en diversos sectores empresariales

RA14. Mat 1.1 Planteará unas hipótesis sobre un conjunto de datos realizando un test de hipótesis que te permita saber si debes aceptar o rechazar la hipótesis nula.

RA14. Mat 1.2 Propondrá un algoritmo para predecir la aceptación, por parte de clientes banco visionarios, de diferentes campañas.

RA14. Mat 1.3 Construirá un algoritmo de clasificación de pacientes sanos vs enfermos.

RA14. Mat 1.4 Diseña un algoritmo para la segmentación de clientes.

RA14. Mat 1.5 Lista todas las variables que deberías tener en cuenta para optimizar los procesos de almacenaje de una empresa.

RA14. Mat 1.6 Formulará nuevas aplicaciones que podría desarrollar en el ámbito empresarial o sectorial utilizando las herramientas más punteras de analítica de datos.

### **3. PLAN DE APRENDIZAJE DE LA ASIGNATURA**

#### **Metodología docente**

La metodología del curso será mixta, con una parte orientada a la transmisión de conceptos básicos y centrada en la exposición de conocimientos, y una segunda parte centrada en el uso de herramientas y realización de casos prácticos partiendo de las mismas que habiliten para el uso de los conceptos teóricos vistos con anterioridad.

**Horas de dedicación (horas lectivas + trabajo del alumno): 75**

**Evaluación (sistema de evaluación, sistema de cualificación...)**

La evaluación se realizará a través de las siguientes actividades:

- (1) Examen teórico-práctico a realizar al acabar el curso, con un valor igual al 40% de la nota final.
- (2) Un ejercicio práctico desarrollado de forma grupal a realizar durante el curso, cada uno con un valor igual al 50% de la nota final.
- (3) Evaluación continuada con un valor igual al 10% de la nota final.

En caso que se suspenda el curso, se realizará un examen teórico-práctico recuperativo que tomará un valor igual al 100% de la nota final.

Las clases tendrán dos o tres componentes:

- (1) Todas las clases tienen componente de clase magistral en que se expondrán los conceptos de la materia, su base teórica, aplicabilidad y limitaciones, estas dos últimas específicamente centradas en ejemplos.
- (2) Todas las clases tienen un componente de exposición práctica de una herramienta de software para la estadística. En el curso se utilizará Python y sus ecosistema de paquetes para el análisis de datos (pandas, statmodels y/o scikit learn).
- (3) En algunas sesiones se propondrán ejercicios abiertos a realizar por el alumnado individualmente o en grupos. Estos ejercicios tendrán un componente de análisis y otro de resolución utilizando las herramientas expuestas en clase. Uno de estos ejercicios es obligatorio, se realizará fuera del horario de las sesiones y contribuyen a la nota final del curso tal como se indica en el sistema de evaluación. La presentación de los mismos se realizará en la tercera sesión donde también se abordarán de forma conjunta y colaborativa diferentes aspectos de los mismos.

\*Información sobre las sesiones.

Primera sesión	Conceptos de estadística y preparación de datos. Toma de contacto con el uso de Python para su uso en estadística.
Segunda sesión	Regresión lineal, caso de estudio.
Tercera sesión	Temas avanzados de regresión, distribuciones estadísticas, muestro, población y estimadores. Presentación del ejercicio de curso.
Cuarta sesión	Intervalos de confianza, test de hipótesis
Quinta sesión	Tipos de errores, Regresión logística, caso de estudio
Sexta sesión	Técnicas de reducción de dimensionalidad y examen final.

#### **4. PROFESORADO**

**Jordi Pereira** es doctor en Administración y Dirección de Empresas, ingeniero en Organización Industrial y en Electrónica Industrial por la Universitat Politècnica de Catalunya. Lleva a cabo actividades de investigación en el diseño de algoritmos para la optimización de problemas de planificación y programación de operaciones en producción y logística. Trabajó en la Universidad Politècnica de Catalunya entre 2003 y 2014, ejerciendo posteriormente en la Universidad Católica del Norte y la Universidad Adolfo Ibáñez en Chile. Desde el 2022 es Senior Lecturer en la UPF-BSM. A lo largo de estos años ha sido miembro de diversos comités de certificación de la calidad docente e

investigadora, ha recibido varios premios de excelencia y ha dirigido y participado en diversos proyectos de investigación con financiación pública y privada.

Publicaciones recientes:

- Álvarez-Miranda, E., Pereira, J., Vilà, M., A Branch, bound and remember algorithm for maximizing the production rate in the assembly line balancing problem, *Computers and Operations Research* 166, 106597, 2024.
- Álvarez-Miranda, E., Pereira, J., Vilà, M., Analysis of the Simple Assembly Line Balancing Problem Complexity, *Computers and Operations Research* 159, 106323, 2024.
- Lagos, F., Pereira, J., Multi-Armed Bandit-Based Hyper-Heuristics for Combinatorial Optimization Problems, *European Journal of Operational Research* 312, 70-91, 2024.
- Pereira, J., Ritt, M., Exact and heuristic methods for a workload allocation problem with chain precedence constraints, *European Journal of Operational Research* 309, 387-398, 2023.
- Álvarez-Miranda, E., Pereira, J., Vargas, C., Vilà, M., Variable-depth local search heuristic for assembly line balancing problem, *International Journal of Production Research* 61, 3103-3121, 2023.
- Álvarez-Miranda, E., Pereira, J., Vargas, C., Vilà, M., Variable-depth local search heuristic for assembly line balancing problem, *International Journal of Production Research* 61, 3103-3121, 2023.

##### 5. **BIBLIOGRAFIA** (obligatoria/recomendada)

- Daniel Peña, Fundamentos de estadística, Alianza Editorial, 2014.
- Peter Bruce, Andrew Bruce, Peter Gedeck, Estadística práctica para ciencia de datos con R y Python, Marcombo, 2022.
- Sheldon M. Ross, Introducción a la estadística, Editorial Reverté, 2014.
- Daniel Peña, Regresión y diseño de experimentos, Alianza Editorial, 2010.