

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

<b>Código</b>	36502
<b>Nombre</b>	Análisis Exploratorio de Datos y Bases de Datos
<b>Ciclo</b>	Grado
<b>Créditos ECTS</b>	6.0
<b>Curso académico</b>	2023 - 2024

**Titulación(es)**

<b>Titulación</b>	<b>Centro</b>	<b>Curso</b>	<b>Periodo</b>
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	Facultad de Economía	1	Primer cuatrimestre

**Materias**

<b>Titulación</b>	<b>Materia</b>	<b>Caracter</b>
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	5 - Fundamentos del Análisis de Datos	Formación Básica

**Coordinación**

<b>Nombre</b>	<b>Departamento</b>
COLL SERRANO, VICENTE	110 - Economía Aplicada

**RESUMEN**

*Análisis Exploratorio de datos y bases de datos* es una asignatura de formación básica adscrita al área de Métodos Cuantitativos para la Economía y la Empresa que se imparte en el primer semestre del primer curso del Grado de Inteligencia y Analítica de Negocios con una carga lectiva total de 6 créditos ECTS.

El análisis exploratorio de datos (E.D.A, Exploratory data analysis), desarrollado por John W. Tukey, es la primera etapa del proceso de análisis de datos y consiste en un conjunto de métodos estadísticos y gráficos que permiten:

- organizar y estructurar los datos.
- explorar la distribución de las variables consideradas con la finalidad, por ejemplo, de detectar la presencia de valores missing, valores anómalos, comprender la dispersión y forma de los datos, etc.
- entender las relaciones entre las variables
- comprobar el cumplimiento de los supuestos en los que se basan gran parte de métodos multivariantes.
- reorganizar y reestructurar los datos para posteriores procedimientos de análisis



Para capacitar al estudiante en el análisis exploratorio de datos, en la asignatura se estudian los conceptos básicos de estadística descriptiva, tanto de carácter unidimensional como multidimensional, y se promueve el trabajo con bases de datos reales utilizando como software de análisis estadístico el programa R y RStudio.

## CONOCIMIENTOS PREVIOS

### Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

### Otros tipos de requisitos

No se requieren.

## COMPETENCIAS

### 1332 - Grado en Inteligencia y Analítica de Negocios/BIA

- 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.
- Capacidad de acceso y gestión de la información en diferentes formatos para su posterior análisis a fin de obtener conocimiento a través de datos.
- Capacidad de análisis y síntesis
- Capacidad para analizar y buscar información proveniente de fuentes diversas.
- Capacidad para utilizar las TIC, tanto en el ámbito de estudio como en el desarrollo profesional.
- Aplicar métodos y técnicas de análisis, síntesis y representación gráfica mediante programas informáticos.
- Conocer las distintas tipologías de datos.
- Reorganizar y reestructurar variables y bases de datos.

## RESULTADOS DE APRENDIZAJE

- Conocer las distintas tipologías de datos con las que se puede trabajar para resolver problemas reales e identificarlas.
- Saber acceder a datos de distinta tipología.
- Aplicar diferentes métodos y técnicas de análisis exploratorio mediante el uso de programas informáticos (R y RStudio).
- Conocer el razonamiento lógico/estratégico para abordar situaciones reales del mundo económico.
- Saber buscar, organizar, seleccionar y valorar información.
- Saber organizar y depurar datos y bases de datos y como fusionarlas.
- Manejar herramientas cuantitativas básicas y su aplicación al entorno económico.



- Incremento de la capacidad para identificar, clasificar, razonar, argumentar e interpretar las relaciones entre las variables económicas.
- Aumento de la habilidad de utilizar el razonamiento lógico/estratégico para abordar situaciones reales del mundo económico.

## DESCRIPCIÓN DE CONTENIDOS

### 1. Tema 1. Introducción al Análisis Exploratorio de Datos

- Tipos de variables. Escalas.
- Distribuciones de frecuencias unidimensionales. Tabulación de datos.
- Resúmenes estadísticos: tendencia, dispersión y forma.
- Detección de valores anómalos y ausentes. Visualización.

### 2. Tema 2. R para el análisis estadístico de datos. Introducción.

- Tipos de datos en R
- Vectores, matrices, listas y dataframes.

### 3. Tema 3. Acceso, limpieza y depuración de datos

- Ficheros de texto: csv, txt
- Formatos propietarios: Excel, SPSS, Stata, etc.
- Formatos de R: RData, rds.
- Acceso a datos a través de API, datos no estructurados, etc.
- Data cleaning: el proceso de limpieza y depuración de datos

### 4. Tema 4. Iniciación a la Visualización de datos.

- Histograma
- Diagrama de barras
- Diagrama de caja (boxplot)
- Gráficos de líneas

### 5. Tema 5. Transformando los datos

- ¿Qué son datos tidy?
- Reshaping los datos: como pasar de formato ancho a formato long y al contrario.
  - o gather()
  - o spread()
  - o separate()
  - o unite()

**6. Tema 6. Análisis de datos multidimensionales.**

- Tablas de frecuencias conjuntas y marginales
- Vector de medidas centrales y matriz de varianzas-covarianzas
- Relación entre variables (covarianza, correlación, etc.)
- Regresión y predicción.
- Valores ausentes e imputación.

**7. Tema 7. Data wrangling: Manejando/gestionando los datos.**

- Gramática para el manejo de datos con R: el paquete dplyr
  - o select()
  - o filter()
  - o group\_by()
  - o summarize()
  - o arrange()
  - o mutate()
- Unir/fusionar bases de datos:
  - o Ideal: r\_bind() y c\_bind()
  - o Mutating joins: inner\_join(), left\_join(), right\_join(), full\_join()

**VOLUMEN DE TRABAJO**

ACTIVIDAD	Horas	% Presencial
Prácticas en aula informática	45,00	100
Clases de teoría	15,00	100
Elaboración de trabajos individuales	10,00	0
Estudio y trabajo autónomo	26,00	0
Preparación de actividades de evaluación	14,00	0
Preparación de clases de teoría	10,00	0
Preparación de clases prácticas y de problemas	20,00	0
Resolución de casos prácticos	10,00	0
<b>TOTAL</b>	<b>150,00</b>	

**METODOLOGÍA DOCENTE**

El desarrollo de la asignatura se estructura fundamentalmente en torno a las sesiones teóricas y las sesiones prácticas. Según el tipo de sesión (teórica o práctica) se elegirá un método didáctico u otro.



En las sesiones teóricas, con una duración de 1 hora, se expondrán los contenidos, de forma teórico-práctica, correspondientes a los fundamentos de estadística descriptiva y del análisis exploratorio de datos.

El método docente predominante en las clases teóricas será la clase magistral participativa. Esta metodología permite dirigir de forma organizada los grupos grandes de estudiantes ofreciendo las ventajas de una clase magistral sin limitar por ello, la participación de los estudiantes y la interacción profesor-estudiante. Se intentará fomentar la participación y la discusión en la clase, con el fin de ofrecer al alumno una implicación directa con el contenido.

En las sesiones prácticas, que tienen una duración de 3 horas, el profesor guiará a los estudiantes en el manejo de R y RStudio y su aplicación al análisis exploratorio de datos a través del uso de bases de datos reales. Para afianzar el conocimiento adquirido en la asignatura, se propondrá a los estudiantes la resolución de casos prácticos similares, pero no totalmente idénticos, a los tratados en clase. Dada la naturaleza de las sesiones prácticas, es indispensable el seguimiento semanal de los contenidos estudiados por parte de los estudiantes.

Las sesiones prácticas se desarrollarán siguiendo distintas estrategias docentes en función de los contenidos discutidos en las sesiones teóricas y prácticas.

## EVALUACIÓN

La evaluación de la asignatura se realizará teniendo en cuenta la evaluación continua y una prueba de síntesis al final del cuatrimestre.

La prueba de síntesis, no necesariamente de carácter escrito, constará de preguntas teórico-prácticas que permitan valorar si el estudiante ha asimilado los conceptos clave del programa. Esta prueba de síntesis supondrá el 70% de la nota final.

La evaluación continua tiene como objetivo desarrollar las competencias de los estudiantes y estimular el trabajo diario. Por tanto, esta parte de la evaluación se basará en una valoración del seguimiento de la materia. La evaluación continua supondrá el 30% de la nota final. Por su propia naturaleza, las pruebas y/o actividades de evaluación continua son **NO RECUPERABLES**.

La nota final será la suma ponderada de la prueba de síntesis y de evaluación continua. En el caso que no se supere la prueba de síntesis (3.5 puntos sobre 7, o alternativamente 5 puntos sobre 10), la nota final no podrá superar el máximo de 4.5 puntos.

El estudiante que no participe de la evaluación continua podrá ser evaluado de la prueba de síntesis y podrá obtener una nota final máxima de 7 puntos. Para superar la asignatura el estudiante tendrá que obtener un mínimo de 5 puntos sobre 7 (o alternativamente, 7 puntos sobre 10) en la prueba de síntesis.

El sistema de evaluación aplica tanto a la primera como a la segunda convocatoria.



## REFERENCIAS

### Básicas

- Baumer, B.S.; Kaplan, D.T. y Horton, N.J. (2017) Modern Data Science with R. CRC Press.
- CEACES, Proyecto (Contenedor Hipermedia de Estadística Aplicada a las Ciencia Económicas y Sociales). Universitat de València. ON LINE: <http://www.uv.es/ceaces>
- Esteban, J. y otros (2006). Estadística Descriptiva y nociones de probabilidad. Paraninfo, Madrid.
- Murgui, S. y otros (2002). Ejercicios de Estadística. Economía y Ciencias Sociales. Valencia: Tirant lo Blanch.
- Peng, R.D. (2015). Exploratory Data Analysis with R. Lean publishing. (bookdown: <https://bookdown.org/rdpeng/exdata/>)
- Teetor, P. (2011). R Cookbook. OREILLY
- Wickham, H. y Golemund, G. (2016). R for Data Science. OREILLY.

### Complementarias

- Newbold, P. y otros (2008). Estadística para Administración y Economía. Madrid Pearson-Prentice Hall, (6ª Edición).
- Pearson, R. K. (2018). Exploratory Data Analysis using R. Chapman and Hall/CRC.