



# Curso Académico 2024-25

## ESTUDIO Y DEPURACIÓN DE DATOS

### Ficha Docente

#### ASIGNATURA

Nombre de asignatura (Código GeA): ESTUDIO Y DEPURACIÓN DE DATOS (801598)

Créditos: 6

Créditos presenciales: 2,40

Créditos no presenciales: 3,60

Semestre: 4

#### PLAN/ES DONDE SE IMPARTE

Titulación: GRADO EN ESTADÍSTICA APLICADA

Plan: GRADO EN ESTADÍSTICA APLICADA

Curso: 2 Ciclo: 1

Carácter: Obligatoria

Duración/es: Segundo cuatrimestre (actas en Jun. y Jul.)

Idioma/s en que se imparte: Español

Módulo/Materia: ANÁLISIS DE DATOS/PREPARACIÓN Y DEPURACIÓN DE DATOS. HERRAMIENTOS DE APOYO

#### PROFESOR COORDINADOR

Nombre	Departamento	Centro	Correo electrónico	Teléfono
PINEDA SANJUAN, SILVIA	Estadística y Ciencia de los Datos	Facultad de Estudios Estadísticos	sipineda@ucm.es	

#### PROFESORADO

Nombre	Departamento	Centro	Correo electrónico	Teléfono
PINEDA SANJUAN, SILVIA	Estadística y Ciencia de los Datos	Facultad de Estudios Estadísticos	sipineda@ucm.es	
GAMBOA PEREZ, MARIA	Estadística y Ciencia de los Datos	Facultad de Estudios Estadísticos	mgamboa@ucm.es	

#### SINOPSIS

##### BREVE DESCRIPTOR:

Depuración, codificación, transformaciones y tratamiento previo al análisis estadístico de una base de datos.

##### REQUISITOS:

Conocimientos básicos de R (se recomienda haber cursado Software II)

Conocimientos básicos de descripción y exploración de datos, azar y probabilidad y estimación I y II.

##### OBJETIVOS:

##### COMPETENCIAS:

##### Generales

CG10 Desarrollar habilidades que favorezcan el trabajo en grupo

##### Transversales:

##### Específicas:

CE2 Preparar y depurar la información disponible para su posterior tratamiento estadístico

CE4 Identificar y organizar la información relevante de un problema

CE21 Utilizar correctamente software estadístico programable y manejar al menos dos de los más utilizados en el mundo empresarial

##### Otras:

##### CONTENIDOS TEMÁTICOS:

TEMA 1. ARCHIVOS DE DATOS: CODIFICACIÓN Y TRANSFORMACIÓN

1.1 Tipos de variables

1.2 Codificación y transformación de datos

1.3 Gráficos y tablas

TEMA 2. CONTROL DE INTEGRIDAD DE LOS DATOS



# Curso Académico 2024-25

## ESTUDIO Y DEPURACIÓN DE DATOS

### Ficha Docente

- 2.1 Detección y tratamiento de datos atípicos (univariante)
- 2.2 Detección y tratamiento de datos atípicos (bivariante)
- 2.3 Detección y tratamiento de datos atípicos (multivariante)
- 2.4 Revisión de datos duplicados

#### TEMA 3. VALORES PERDIDOS: DETECCIÓN Y TRATAMIENTO

- 3.1 Detección de datos perdidos
- 3.2 Tratamiento de datos perdidos (imputación simple)
- 3.3 Tratamiento de datos perdidos (imputación múltiple)

#### TEMA 4. EVALUACIÓN DE LAS HIPÓTESIS DE PARTIDA PARA EL ANÁLISIS DE DATOS

- 4.1 Pruebas de normalidad
- 4.2 Pruebas de homocedasticidad

#### ACTIVIDADES DOCENTES:

##### Clases teóricas:

50%

##### Seminarios:

##### Clases prácticas:

50%

##### Trabajos de campo:

##### Prácticas clínicas:

##### Laboratorios:

##### Exposiciones:

##### Presentaciones:

##### Otras actividades:

##### TOTAL:

100%

##### EVALUACIÓN:

Se valorará la nota final a través de los conocimientos adquiridos mediante el desarrollo de ejercicios, trabajos y participación en el aula, que contribuirán con el 40% a la nota final. El porcentaje restante, el 60%, corresponderá al examen final.

La nota final tendrá en cuenta tanto la evaluación continua como la prueba final y se calculará como el máximo entre:

- a) Calificación de la prueba final.
- b) La media ponderada de la evaluación continua y la prueba final, siendo el peso de la evaluación continua del 40%.

Cualquier alumno podrá presentarse al examen final, siendo la valoración del mismo el 100% de su nota final.

En todo caso, sí se podrá superar la asignatura mediante el procedimiento de evaluación continua.

##### BIBLIOGRAFÍA BÁSICA:

Introducción a la ciencia de datos con R: Preparación de los datos y análisis no supervisado. Aida Calviño Martínez, Juana María Alonso Revenga. García Maroto Editores, 2022

Alonso Revenga, Juana M. "Depuración de Datos con SPSS y SAS" García Maroto Editores, 2019

Boehmke, Bradley. "Data Wrangling with R". Springer. 2016

Cody, Ron. "Cody's Data Cleaning Techniques Using SAS Software". Ed SAS Publishin; 1999

López, Ana M. "Análisis previo y exploratorio de datos", [en línea], <https://personal.us.es/analopez/aed.pdf>

Salvador Figueras, M y Gargallo, P. (2003): "Análisis Exploratorio de Datos", [en línea]



# Curso Académico 2024-25

## ESTUDIO Y DEPURACIÓN DE DATOS

### Ficha Docente

<http://www.5campus.com/leccion/aed>

R for data science: import, tidy, transform, visualize, and model data. Wickham, H., Golemund, G. (2017): <https://r4ds.hadley.nz/>

Data Wrangling with R. Boehmke, Bradley. Springer. 2016

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>

#### **OTRA INFORMACIÓN RELEVANTE**