



Curso Académico 2024-25

ESTUDIO Y DEPURACIÓN DE DATOS

Ficha Docente

ASIGNATURA

Nombre de asignatura (Código GeA): ESTUDIO Y DEPURACIÓN DE DATOS (806310)

Créditos: 6

Créditos presenciales: 2,40

Créditos no presenciales: 3,60

Semestre: 4

PLAN/ES DONDE SE IMPARTE

Titulación: GRADO EN CIENCIA DE LOS DATOS APLICADA

Plan: GRADO EN CIENCIA DE LOS DATOS APLICADA

Curso: 2 **Ciclo:** 1

Carácter: Obligatoria

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

Idioma/s en que se imparte:

Módulo/Materia: /

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	

SINOPSIS

BREVE DESCRIPTOR:

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

REQUISITOS:

Se recomienda tener conocimientos de R o haber cursado Software 2 y estar cursando análisis supervisado I

OBJETIVOS:

COMPETENCIAS:

Generales

CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

CG10 - Desarrollar la capacidad de expresar y aplicar rigurosamente los conocimientos adquiridos en la resolución de problemas.

CG11 - Poner en práctica las técnicas, herramientas y metodologías propias de la Ciencia de los Datos en distintos ámbitos de aplicación.

Transversales:

Específicas:

CE2 - Identificar y seleccionar las principales fuentes de información e indicadores en diferentes ámbitos de aplicación como los sociales, económicos, bibliométricos, sanitarios, financieros, geográficos, etc. , incluyendo los aspectos de protección de datos.

CE4 - Identificar la utilidad y el potencial de las técnicas estadísticas y de análisis de datos adquiridas en las distintas áreas de aplicación (como la documental, geográfica, actuarial, económica, financiera, biosanitaria, etc.) y utilizarlas adecuadamente para descubrir patrones de comportamiento en datos de cualquier volumen y tipología y extraer conclusiones relevantes.

CE12 - Identificar y aplicar las técnicas adecuadas de exploración, gestión, procesamiento, explotación y visualización de datos en función de su volumen y complejidad.

CE13 - Identificar y aplicar las técnicas para la extracción de información, preparar y depurar la información disponible para su posterior análisis de datos.

Otras:



Curso Académico 2024-25

ESTUDIO Y DEPURACIÓN DE DATOS

Ficha Docente

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

- 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% y de la prueba final del 60%.

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

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

Depuración de datos con SPSS y SAS. Juana María Alonso Revenga. García Maroto Editores, 2019



Curso Académico 2024-25

ESTUDIO Y DEPURACIÓN DE DATOS

Ficha Docente

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

OTRA INFORMACIÓN RELEVANTE