

**Proyecto/Guía docente de la asignatura**

<b>Asignatura</b>	Análisis de Datos		
<b>Materia</b>	2- Técnicas Estadísticas		
<b>Módulo</b>			
<b>Titulación</b>	Grado en Estadística		
<b>Plan</b>	549	<b>Código</b>	47093
<b>Periodo de impartición</b>	Cuatrimestre 1º	<b>Tipo/Carácter</b>	OB
<b>Nivel/Ciclo</b>	Grado	<b>Curso</b>	3
<b>Créditos ECTS</b>	6		
<b>Lengua en que se imparte</b>	Castellano. Se manejarán materiales teóricos y programas informáticos en inglés.		
<b>Profesor/es responsable/s</b>	Luis Ángel García y Miguel Alejandro Fernández		
<b>Datos de contacto (E-mail, teléfono...)</b>	lagarcia@eio.uva.es, miguelaf@eio.uva.es		
<b>Departamento</b>	Estadística e Investigación Operativa		

#1# ver la memoria de verificación del grado en <http://www.eio.uva.es/docencia/grado/memoria.pdf>

**1. Situación / Sentido de la Asignatura #1#****1.1 Contextualización #1#**

Esta asignatura obligatoria de Análisis de Datos de Tercer Curso del Grado en Estadística constituye el primer contacto del estudiante con problemas de estadística multivariante y el análisis de grandes tablas de datos. Con ella se busca iniciar al estudiante en los problemas multivariantes y sus técnicas, tratados desde una perspectiva fundamentalmente descriptiva.

Tiene un enfoque eminentemente práctico, pero siempre tratando de comprender en profundidad las ideas que se manejan y las bases conceptuales de cada técnica, evitando presentar al alumno un catálogo exhaustivo de procedimientos para que los utilice mediante "recetas" mecánicamente aplicadas.

**1.2 Relación con otras materias #1#**

Aunque la amplitud de la Estadística obliga a distribuir en asignaturas específicas el estudio de diferentes métodos, esta asignatura ayudará a proporcionar al estudiante una visión más homogénea y global, contribuyendo a la adquisición de una parte del bagaje de "pensamiento estadístico" que un buen profesional debe utilizar.

Se abordarán diversos problemas completos desde su planteamiento en la vida real, pasando por la manipulación de datos, su análisis y la aplicación práctica de los procedimientos estadísticos multivariantes más apropiados, hasta la elección de una solución final satisfactoria y su presentación mediante la elaboración de informes y su exposición oral. Los estudiantes tratarán problemas complejos que deberán estructurar en



tareas más simples, fraccionándolos en pasos o etapas encadenadas y relacionadas. Se potenciará el uso de diversos programas informáticos estadísticos (SAS, R...) y otros auxiliares que empleará como herramienta intermedia para preparación, lectura y depuración de los datos iniciales y para elaborar los informes finales. Deberá también utilizar técnicas estadísticas diversas aprendidas en otras asignaturas del Grado, como Regresión y Anova o Inferencia Estadística. Con ello se persigue que el alumno progrese en su proceso de maduración dentro de su formación estadística integral.

### 1.3 Prerrequisitos #1#

---

Para seguir la asignatura es conveniente por parte del alumno un manejo desenvuelto, conceptual y práctico, de Estadística Descriptiva, Cálculo de Probabilidades e Inferencia Estadística. También se precisan conocimientos de Álgebra y Geometría y de Cálculo Infinitesimal, así como de los elementos básicos de los Lenguajes de Programación, todo ello al nivel que se imparte en los cursos 1º y 2º del Grado. También se manejarán algunas ideas propias de los Modelos Lineales, que se desarrollan de forma paralela en asignaturas obligatorias de 3º del Grado. Deberá leer inglés técnico y manejar programas estadísticos en inglés.

## 2. Competencias #1#

---

Se incluyen a continuación las competencias de la asignatura, según aparecen en la memoria de verificación del Grado:

### Transversales Instrumentales

- I1. Capacidad de análisis y síntesis
- I2. Capacidad de gestión de la información
- I3. Capacidad de organización y planificación
- I4. Conocimientos de informática relativos al ámbito de estudio
- I5. Resolución de problemas
- I6. Comunicación oral y escrita en lengua nativa
- I7. Conocimiento de lenguas extranjeras
- I8. Toma de decisiones

### Personales

- P2. Razonamiento crítico
- P4. Compromiso ético

### Sistémicas

- S1. Aprendizaje autónomo
- S2. Adaptación a nuevas situaciones
- S3. Motivación por el trabajo bien hecho
- S4. Iniciativa y espíritu emprendedor
- S5. Creatividad

### 2.1 Generales #1#

---

G1. Capacidad para la gestión de la información: Incluye la capacidad para la búsqueda, manejo y exposición de información relevante de diversas fuentes, así como el manejo de las herramientas TIC necesarias.



G2. Capacidad para la abstracción y el razonamiento crítico: El modelado y análisis de datos de problemas reales exige una elevada capacidad de abstracción, y el razonamiento crítico es necesario para hacer interpretaciones y establecer conclusiones y soluciones con rigor científico.

G3. Capacidad para la puesta al día y el auto-aprendizaje: Incluye la capacidad para la búsqueda de recursos que permitan la solución de nuevos problemas o de nuevas técnicas, en un medio científico y tecnológico en continua evolución.

## 2.2 Específicas #1#

E1. Recogida y tratamiento de datos: Incluye la capacidad para decidir sobre el diseño del procedimiento de obtención de datos. Capacidad para la búsqueda de información de fuentes diversas y para la elaboración de cuestionarios. Dominio en el manejo de bases de datos y en el tratamiento y depuración de los mismos.

E2. Descripción y síntesis de datos: Esta competencia es la que permitirá describir numéricamente y mediante gráficos diferentes características de interés de variables e individuos de los datos objeto de estudio, localización, escala, diferentes tipos de asociación, outliers... con empleo y adaptación de la herramienta informática apropiada o programación específica.

E3. Ajuste de modelos estadísticos y de investigación operativa: Incluye la capacidad para crear o reconocer un modelo adecuado al problema objeto de estudio. Capacidad para el ajuste del modelo mediante las herramientas estadísticas e informáticas adecuadas.

E4. Análisis de resultados, interpretación y validación de modelos: Incluye la capacidad para la selección y validación de un modelo. Capacidad para la interpretación de los modelos ajustados y las diferencias entre ellos. Empleo de herramientas informáticas específicas.

E5. Extracción de conclusiones: Incluye la capacidad para interpretar los resultados del ajuste de un modelo seleccionado en términos del problema objeto de estudio, evaluando su utilidad y/o proponiendo la necesidad de otras orientaciones del estudio.

E6. Presentación y comunicación de resultados: Incluye la capacidad para presentar los resultados de los análisis realizados, junto a las posibles soluciones a los problemas planteados por los demandantes del estudio en contextos diversos.

E7. Dominio de conceptos y herramientas informáticas a nivel de usuario avanzado: programación, sistemas operativos, algoritmos, computación, inteligencia artificial, aprendizaje automático, almacenes y minería de datos, etc.

## 3. Objetivos #1#

Objetivos generales prioritarios:

- Conocimiento teórico y práctico de técnicas descriptivas de análisis de datos multivariante.
- Utilización de diversos paquetes estadísticos para estas técnicas como R y SAS

## 4. Contenidos y/o Bloques temáticos

### Bloque 1: ANÁLISIS DE DATOS

Carga de trabajo en créditos ECTS:



**a. Contextualización y justificación #1#**

---

Esta asignatura obligatoria de Análisis de Datos de Tercer Curso del Grado en Estadística constituye el primer contacto del estudiante con problemas de estadística multivariante y el análisis de grandes tablas de datos. Con ella se busca iniciar al estudiante en los problemas multivariantes y sus técnicas.

Tiene un enfoque eminentemente práctico, pero siempre tratando de comprender bien las bases conceptuales de cada técnica y evitando presentar al alumno un catálogo exhaustivo de procedimientos para que los utilice mediante "recetas" mecánicamente aplicadas.

Se tratará de dotar al estudiante de la herramienta estadística y geométrica formal mínima, más a nivel de ideas que de desarrollos teóricos, para que comprenda el funcionamiento de las técnicas estadísticas que maneja y sea capaz de plantearlas correctamente y realizar interpretaciones precisas.

La asignatura utiliza técnicas descriptivas avanzadas basadas en proyecciones, con gran profusión de gráficos y tablas que el alumno aprenderá a elaborar e interpretar, lo cual requiere un manejo desenvuelto de la geometría en  $R^n$ .

**b. Objetivos de aprendizaje #1#**

---

- Conocimiento teórico y práctico de técnicas descriptivas de análisis de datos multivariante.
- Utilización de diversos paquetes estadísticos para estas técnicas como R y SAS.

**c. Contenidos #1#**

---

- Análisis de Componentes Principales.
- Análisis de Correspondencias.
- Clasificación supervisada. Análisis discriminante.
- Clasificación no supervisada. Análisis clúster.

**d. Métodos docentes #1#**

---

Se combinarán clases magistrales, prácticas en salas informáticas, elaboración, exposición y discusión de trabajos individuales y en grupo.

**e. Plan de trabajo #1#**

---

Se abordarán diversos problemas completos desde su planteamiento en la vida real, pasando por la manipulación de datos, su análisis y la aplicación práctica de los procedimientos estadísticos multivariantes más apropiados, hasta la elección de una solución final satisfactoria y su presentación mediante la elaboración de informes y su exposición oral. Los estudiantes tratarán problemas complejos que deberán estructurar en tareas, fraccionándolos en pasos o etapas encadenadas y relacionadas. Se potenciará el uso de diversos programas informáticos estadísticos (SAS, R) y otros auxiliares que empleará como herramienta intermedia para preparación, lectura y depuración de los datos iniciales, y para elaborar los informes finales. Deberá también utilizar técnicas estadísticas diversas aprendidas en otras asignaturas del Grado, como Regresión y Anova, Modelos Lineales o Inferencia Estadística. Con ello se persigue que el alumno progrese en su proceso de maduración dentro de su formación estadística integral.

**f. Evaluación #1#**

La evaluación de esta asignatura, tanto en la convocatoria ordinaria como en la extraordinaria, se efectuará a partir de las siguientes actividades.

ACTIVIDAD	PESO EN LA NOTA FINAL	OBSERVACIONES
Evaluación continua (podrá incluir prácticas con ordenador evaluadas y trabajos individuales o en grupo, entre otros)	30%	Las notas obtenidas en evaluación continua son válidas para la convocatoria ordinaria.
Examen escrito en la convocatoria ordinaria	70%	
Examen en la convocatoria extraordinaria	100%	Este examen podrá incluir programación con ordenador.

**g. Bibliografía básica #1#**

- FLURY BERNARD (1997). A first Course in Multivariate Statistics. Ed: Springer
- LEBART, L., MORINEAU, A. Y WARWICK, K. M. (1984), Multivariate Descriptive Statistical Analysis, Ed: Wiley.
- PEÑA D. (2002) Análisis de Datos Multivariantes. Ed.: Mc Graw Hill.
- SRIVASTAVA, m. s. (2002) Methods of Multivariate Statistics. Ed. Wiley

**h. Bibliografía complementaria #1#**

- AFIFI, A. A. Y CLARK, V. (1990), Computer-Aided Multivariate Analysis (second edition), Ed: Van Nostrand Reinhold.
- ALUJA, T Y MORINEAU, A (1999) Aprender de los Datos: El Análisis de Componentes Principales. Una aproximación desde el Data Mining. Ed: EUB S.L.
- CHATFIELD, C. Y COLLINS, A.J. (1980), Introduction to Multivariate Analysis, Ed: Chapman and Hall.
- EVERITT, B. (1993), Cluster Analysis (third edition), Ed: Edward Arnold.
- EVERITT B. S. Y DUNN G. (1991), Applied Multivariate Data Analysis, Ed: Edward Arnold, London.
- GREENACRE, M. J. (1984). Theory and Applications of Correspondence Analysis. Academic Press.
- JOBSON, D.V. (1992). Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods. Ed: Springer-Verlag.
- KRZANOWSKI, W. J. (1988), Principles of Multivariate Analysis (a user's perspective), Ed: Oxford Science Publications.
- LEBART, L., MORINEAU, A. y FENELON J. (1985) Tratamiento estadístico de datos. Ed.: Marcombo.

**i. Recursos necesarios #1#**

Software estadístico específico, disponible en las salas de informática de la Facultad de Ciencias y el Departamento de Estadística e Investigación Operativa.



**j. Temporalización #1#**

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
ANÁLISIS DE DATOS	6	Primer Cuatrimestre

**5. Métodos docentes y principios metodológicos #1#**

MÉTODOS DOCENTES	OBSERVACIONES
Clases expositivas y de problemas	30 horas
Laboratorios	30 horas
Tutorías	

La modalidad docente prevista es la de PRESENCIALIDAD SEGURA, con un grado de presencialidad del 40%. Es decir, si la disponibilidad de espacios en el centro y la organización global prevista en el mismo lo permite, la docencia se desarrollará de forma presencial. En otro caso, la docencia será bimodal con webcam en aulas para que los alumnos que no puedan seguir la docencia de forma presencial la sigan en streaming preferentemente desde un espacio contiguo.

**6. Tabla de dedicación del estudiante a la asignatura #1#**

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teóricas	30	Estudio y trabajo autónomo	30
Laboratorios/Clases prácticas	30	Elaboración de trabajos	20
		Trabajo personal en laboratorio	25
		Preparación de presentaciones	10
		Otras actividades complementarias	5
<b>Total presencial</b>	<b>60</b>	<b>Total no presencial</b>	<b>90</b>

**7. Sistema y características de la evaluación #1#**

La calificación de esta asignatura, tanto en la convocatoria ordinaria como en la extraordinaria, se efectuará a partir de los siguientes instrumentos/procedimientos.

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Evaluación continua (podrá incluir prácticas con ordenador evaluadas y trabajos individuales o en grupo, entre otros)	30%	Las notas obtenidas en evaluación continua son válidas para la convocatoria ordinaria.
Examen en la convocatoria ordinaria	70%	
Examen en la convocatoria extraordinaria	100%	Este examen podrá incluir programación con ordenador