



## DATOS IDENTIFICATIVOS

### Análisis exploratorio de datos

Asignatura	Análisis exploratorio de datos			
Código	V03M184V01101			
Titulación	Máster Universitario en Técnicas Estadísticas			
Descriptores	Creditos ECTS	Seleccione	Curso	Cuatrimestre
	5	OP	1	1c
Lengua Impartición				
Departamento	Estadística e investigación operativa			
Coordinador/a	Roca Pardiñas, Javier			
Profesorado	Roca Pardiñas, Javier Sestelo Pérez, Marta			
Correo-e	roca@uvigo.es			
Web	<a href="http://eio.usc.es/pub/mte/">http://eio.usc.es/pub/mte/</a>			
Descripción general				

## Resultados de Formación y Aprendizaje

Código	
C1	Conocer, identificar, modelar, estudiar y resolver problemas complejos de estadística e investigación operativa, en un contexto científico, tecnológico o profesional, surgidos en aplicaciones reales.
C2	Desarrollar autonomía para la resolución práctica de problemas complejos surgidos en aplicaciones reales y para la interpretación de los resultados de cara a la ayuda en la toma de decisiones.
C6	Adquirir conocimientos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas específicamente a la ayuda en la toma de decisiones, y desarrollar la capacidad de reflexión para evaluar y decidir entre distintas perspectivas en contextos complejos.
C8	Adquirir conocimientos teórico-prácticos avanzados de las técnicas destinadas a la realización de inferencias y contrastes relativos a variables y parámetros de un modelo estadístico, y saber aplicarlos con autonomía suficiente un contexto científico, tecnológico o profesional.
C9	Conocer y saber aplicar con autonomía en contextos científicos, tecnológicos o profesionales, técnicas de aprendizaje automático y técnicas de análisis de datos de alta dimensión (big data).
C10	Adquirir conocimientos avanzados sobre metodologías para la obtención y el tratamiento de datos desde distintas fuentes, como encuestas, internet, o entornos "en la nube".

## Resultados previstos en la materia

Resultados previstos en la materia	Resultados de Formación y Aprendizaje
Comprender la finalidad, resultados y beneficios del análisis de un conjunto de datos, así como de sus requerimientos, con el fin de permitir una mejor modelización de problemas y experimentos.	C1 C2
Descubrir la problemática del análisis de un conjunto de datos, con el fin de motivar conceptos y técnicas de la teoría de la probabilidad y de la estadística matemática.	C6 C8 C9 C10
Conocer los conceptos, y técnicas numéricas y gráficas, esenciales para el descubrimiento y entendimiento de las estructuras y relaciones contenidas en un conjunto de datos	C1 C2 C6
Adquirir una visión general de algunos de los principales problemas estadísticos y de los grupos de técnicas más apropiadas para resolverlos.	C1 C2 C6

## Contenidos

Tema	
Manejo del software estadístico R.	Descarga e instalación. Lectura y tratamiento de datos. Procedimientos gráficos. Introducción a la programación.
Naturaleza y tipo de datos.	Población y muestra. Tipos de muestreo. Variables aleatorias discretas y continuas.
Medidas descriptivas numéricas.	Media, desviación típica, varianza, mediana, rango, y cuantiles.
Gráficos estadísticos	Principales gráficos estadísticos: gráfico de puntos, gráfico de barras, gráfico de sectores, histogramas y gráficos de densidad.
Tablas de frecuencias.	Construcción de tablas de frecuencias para uno y varios factores. Frecuencias absolutas, relativas y marginales. Construcción de tablas con R.
Estudio de correlación	Definición de covarianza y correlación. Interpretación y representación gráfica. Tipos de correlación. Matrices y gráficos de correlación multivariantes.

## Planificación

	Horas en clase	Horas fuera de clase	Horas totales
Seminario	0	12	12
Prácticas con apoyo de las TIC	15	15	30
Lección magistral	25	23	48
Resolución de problemas y/o ejercicios	10	10	20
Práctica de laboratorio	0	15	15

\*Los datos que aparecen en la tabla de planificación son de carácter orientativo, considerando la heterogeneidad de alumnado

## Metodologías

	Descripción
Seminario	Se mantendrá un servicio de tutoría en grupo a los alumnos. Los alumnos también podrán consultar sus dudas por correo electrónico
Prácticas con apoyo de las TIC	La docencia se desarrollará mediante la resolución de problemas reales o simulados utilizando los modelos tratados en las sesiones magistrales. Se utilizará principalmente el software R.
Lección magistral	La docencia se desarrollará mediante la exposición por parte del profesor de las diferentes técnicas de Análisis Exploratorio de Datos Para ello, los alumnos dispondrán de apuntes elaborados que servirán de material básico para el estudio y en su defecto de material e información sobre bibliografía específica disponible en la biblioteca o en internet.

## Atención personalizada

### Metodologías Descripción

Seminario	Las dudas de los alumnos serán resueltas de manera individual en el horario de *tutorías de los profesores de la materia. También se contemplan *tutorías en grupo.
-----------	---

## Evaluación

	Descripción	Calificación	Resultados de Formación y Aprendizaje
Resolución de problemas y/o ejercicios	Realización de un examen final teórico y/o práctico	60	C1 C6 C8
Práctica de laboratorio	Realización de cuestionarios y actividades que serán planteadas durante el periodo de docencia da materia.	40	C1 C2 C8 C10

## Otros comentarios sobre la Evaluación

Los criterios de evaluación abarcarán el conocimiento teórico y la competencia práctica sobre los contenidos de la asignatura. El alumnado podrá elegir ser evaluado mediante el sistema de Evaluación Continua (EC), o alternativamente optar por una prueba de Evaluación Global (EG). La evaluación por defecto es la EC. El alumnado podrá elegir EG según el procedimiento y el plazo establecido por el centro. La elección de EG supone la renuncia al derecho de seguir evaluándose

mediante las actividades de EC que resten y a la calificación obtenida hasta ese momento en cualquiera de las pruebas que ya se hayan realizado.

Evaluación continua (EC). En caso de no renunciar a la evaluación continua:

- Evaluación: El alumnado tendrá cuatro pruebas de evaluación continua a lo largo del cuatrimestre (resolución de problemas, ejercicios y/o trabajos) con los pesos sobre la calificación final que se indican: la primera (EC1) y segunda prueba (EC2) tendrán un peso de un 25% cada una, la tercera prueba (EC3) tendrá un peso de un 40% y la última prueba (EC5) tendrá un peso de un 10%. En EC1 y EC2 se evaluarán los temas del 1 al 4 (Parte 1 de la materia) mientras que en EC3 y EC4 se evaluarán los temas del 5 al 8 (Parte 2 de la materia).

- No habrá una nota mínima en cada una de las pruebas.

- La nota mínima para aprobar será de 5 puntos.

- Se considerará que el alumnado se ha presentado a la evaluación continua cuando se haya presentado a alguna de las pruebas/actividades de evaluación de esta modalidad.

Evaluación Global (EG) en la primera oportunidad (convocatoria de febrero): El alumnado será evaluado mediante un examen que constará de varias cuestiones teórico-prácticas sobre los contenidos de la materia, y de varias tareas prácticas relacionadas con datos reales y/o simulados. Este examen supondrá el 100% de la nota. La nota mínima para aprobar será de 5 puntos.

Evaluación Global (EG) en la segunda oportunidad (convocatoria extraordinaria de julio): El alumnado será evaluado mediante un examen que constará de varias cuestiones teórico-prácticas sobre los contenidos de la materia, y de varias tareas prácticas relacionadas con datos reales y/o simulados. Este examen supondrá el 100% de la nota. La nota mínima para aprobar será de 5 puntos.

Los estudiantes que que no participen en la EC ni en la EG, figurarán en las actas como "no presentados".

Con las distintas pruebas que se propondrán a lo largo del curso (y la evaluación global), se valorará el nivel de adquisición de las competencias generales CG1-CG5, así como las competencias transversales CT1, CT3 y CT4, y específicas E1, E2, E6 y E8.

---

## **Fuentes de información**

### **Bibliografía Básica**

### **Bibliografía Complementaria**

Crawley, M. J., **Statistics: an introduction using R**, John Wiley and Sons, 2005

Crawley, M.J., **The R book**, John Wiley and Sons, 2013

Devore, Jay L., **Probability and statistics for engineering and sciences**, Cengage Learning, 2015

James, G., Witten, D., Hastie, T. and Tibshirani, R., **An Introduction to Statistical Learning with Applications in R**, Springer, 2013

Kabacoff, R., **R in Action: Data Analysis and Graphics with R**, Manning Publications, 2011

Maindonald, J. H., **Data analysis and graphics using R: an example-based approach**, Cambridge University Press., 2007

Tukey, J.W., **Exploratory Data Analysis**, Addison-Wesley, 1977

Zumel, N., Mount, J., **Practical Data Science with R**, Manning Publications, 2014

Zuur, A., Ieno, A., Meesters, E., **A Beginners` s Guide to R**, Springer, 2009

---

## **Recomendaciones**

### **Otros comentarios**

No se necesita haber cursado ninguna otra asignatura del máster. Sin embargo es fundamental la asistencia regular a las clases para la superación de esta materia, ya que es muy importante el seguimiento del trabajo realizado en el aula.

Los requisitos básicos de esta materia son un conocimiento básico de la Estadística y conocimientos a nivel usuario de Windows. Como ya se ha comentado se utilizará el software libre R.

En algunas sesiones se realizarán cuestionarios y ejercicios sobre temas impartidos para la valoración de la evolución y comprensión de los alumnos sobre la materia.