



UNIVERSIDADE DA CORUÑA *Universidade* deVigo

## PROGRAMA OFICIAL DE POSTGRADO EN ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

### DATOS IDENTIFICATIVOS DE LA MATERIA

Código de la materia:	V03M017V01102
Nombre de la materia:	<b>Análisis Exploratorio de Datos</b>
Número de créditos:	5 ECTS
Curso académico:	2021/2022
Profesorado:	Javier Roca Pardiñas (UVigo) (coordinador de la materia)

### OBJETIVOS Y COMPETENCIAS DE LA MATERIA

1. Saber transmitir de forma clara y precisa, gráfica y numéricamente, la información disponible en una muestra, tanto a un público especializado como no especializado.
2. Comprender y analizar la asociación entre un conjunto de variables.
3. Interpretar correctamente datos complejos surgidos de aplicaciones reales.
4. Manejar de forma autónoma el software necesario para el análisis estadístico de problemas complejos.

### CONTENIDOS DE LA MATERIA

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

## OBSERVACIONES

El desarrollo de los contenidos de la materia se realizará teniendo en cuenta que las competencias a adquirir por el alumnado deben cumplir con el nivel MECES3. De este modo, el estudio de las técnicas exploratorias de datos que se presenten no se limitará a su descripción y uso, sino que se incidirá en su construcción, interpretación, posibles extensiones y limitaciones prácticas en contextos complejos. Se analizarán ejemplos concretos de contextos aplicados fuera del ámbito de la estadística, incidiendo en el carácter multidisciplinar de la formación a adquirir.

## BIBLIOGRAFÍA

- Crawley, M. J. (2005). Statistics: an introduction using R. Ed. John Wiley and Sons.
- Crawley, M.J. (2013) The R book. John Wiley and Sons.
- Devore, Jay L. (2015) Probability and statistics for engineering and sciences,
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.
- Kabacoff, R. (2011). R in Action: Data Analysis and Graphics with R. Manning Publications
- Maindonald, J. H. (2007). Data analysis and graphics using R: an example-based approach. Cambridge University Press.
- Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley.
- Zumel, N., Mount, J. (2014). Practical Data Science with R. Manning Publications
- Zuur, A., Ieno, A., Meesters, E. (2009) A Beginners's Guide to R. Springer

## COMPETENCIAS

En esta materia se trabajarán las competencias básicas, generales y transversales recogidas en la memoria del título. Se indican a continuación cuáles son las competencias específicas, que se potenciarán en esta materia:

- E1 - Conocer, identificar, modelar, estudiar y resolver problemas complejos de estadística e investigación operativa, en un contexto científico, tecnológico o profesional, surgidos en aplicaciones reales.
- E2 - 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.
- E6 - Adquirir conocimientos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas específicamente a la ayuda en la toma de decisiones, y desarrollar capacidad de reflexión para evaluar y decidir entre distintas perspectivas en contextos complejos.
- E8 - 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.
- E9 - 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).
- E10 - Adquirir conocimientos avanzados sobre metodologías para la obtención y el tratamiento de datos desde distintas fuentes, como encuestas, internet, o entornos "en la nube".

## **METODOLOGÍA DOCENTE: ACTIVIDADES DE APRENDIZAJE Y SU VALORACIÓN EN CRÉDITOS ECTS**

**Sesión 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.

**Prácticas de laboratorio:** 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.

**Tutoría en grupo:** Se mantendrá un servicio de tutoría en grupo a los alumnos. Los alumnos también podrán consultar sus dudas por correo electrónico.

## **CRITERIOS Y MÉTODOS DE EVALUACIÓN**

Los criterios de evaluación abarcarán el conocimiento teórico y la competencia práctica sobre los contenidos de la asignatura.

**Evaluación continua (40%):** la evaluación continua se realizará en base a la resolución de problemas, ejercicios y/o trabajos. En particular, se realizarán actividades que serán planteadas durante el periodo de docencia de la materia. La calificación obtenida se conservará entre las oportunidades (ordinaria y extraordinaria) dentro de la convocatoria de cada curso.

Con las distintas actividades que se propondrán a lo largo del curso, se valorará (junto con el examen final) el nivel de adquisición de las competencias generales CG1-CG5, así como las competencias transversales CT1, CT3 y CT4, y específicas E1, E2 y E6.

**Examen final (60%):** El examen final constará de varias cuestiones teórico-prácticas sobre los contenidos de la materia, y de varias tareas prácticas relacionadas con datos reales y/o simulados. En el examen, se evaluarán junto con la parte de evaluación continua, las competencias específicas E1, E2, E6, y E8.

Para superar la materia será necesario obtener una calificación mínima de 4 puntos (sobre 10) en el examen final.

En la segunda oportunidad de evaluación (recuperación) se realizará un examen, y se mantendrá la nota y el peso de la evaluación continua obtenida en la convocatoria ordinaria. La nota final será el máximo de dos cantidades: la nota del nuevo examen y la media ponderada del nuevo examen y la evaluación continua. Para efectuar la ponderación será necesario obtener un mínimo de 4 puntos en el examen.

## **TIEMPO DE ESTUDIO Y DE TRABAJO PERSONAL QUE DEBE DEDICAR UN ESTUDIANTE PARA SUPERAR LA MATERIA**

Se considera que el tiempo de trabajo personal del alumnado para superar la materia es de 122.5 horas repartidas como sigue:

- Actividad presencial: 35 horas (5 ECTS x 7 horas)
- Estudio del material: 52.5 horas (se estima una hora y media por cada hora de docencia presencial)
- Trabajos de evaluación continua: 30 horas

## RECOMENDACIONES PARA EL ESTUDIO DE LA MATERIA

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.

## RECURSOS PARA EL APRENDIZAJE

Bibliografía y apuntes. Uso del campus virtual del Máster, repositorio de material del programa.

## OBSERVACIONES

Para los casos de realización fraudulenta de ejercicios o pruebas, será de aplicación lo recogido en las respectivas normativas de las universidades participantes en el Máster en Técnicas Estadísticas.

Esta guía y los criterios y metodologías en ella descritos están sujetos a las modificaciones que se deriven de normativas y directrices de las universidades participantes en el Máster en Técnicas Estadísticas.

**COVID 19:** En el caso de que sea necesario importar la enseñanza en un modo no presencial, la actividad docente se impartirá mediante alguna plataforma de docencia virtual (Teams, Campus Remoto, etc.) En cualquier caso, todo el material didáctico de la asignatura se pondrá a disposición de los alumnos mediante alguna plataforma de intercambio de información (DropBox, plataforma de tele enseñanza de Faitic, etc.). Del mismo modo, en esta situación, el examen se realizará de forma no presencial, y no se requiere ningún cambio en la planificación docente de esta materia. Además, las tutorías se podrán realizar por medios telemáticos (correo electrónico, videoconferencia, ...) bajo la modalidad de cita previa.