Guia docente					
Datos Identificativos					2022/23
Asignatura (*)	Análisis Estadístico de Datos Complejos			Código	614G02031
Titulación	Grao en Ciencia e Enxeñaría de Datos				
Descriptores					
Ciclo	Periodo	Curso		Tipo	Créditos
Grado	1º cuatrimestre	Cuarto		Optativa	6
Idioma	CastellanoGallego				
Modalidad docente	Presencial				
Prerrequisitos					
Departamento					
Coordinador/a	López Cheda, Ana		Correo electrónico	ana.lopez.cheda@udc.es	
Profesorado	López Cheda, Ana		Correo electrónico	ana.lopez.cheda@udc.es	
Web	https://dm.udc.es/modes/				
Descripción general	Esta materia proporciona un primer contacto del alumnado con las principales técnicas estadísticas para analizar problemas con datos faltantes, datos funcionales, datos censurados o datos sesgados. Se estudiarán los principales mecanismos que provocan la falta de información y se aplicarán las técnicas presentadas a conjuntos de datos reales o simulados. Se analizarán las limitaciones de cada metodología y se realizará el diagnóstico e interpretación de los resultados en términos del problema propuesto.				

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A3	CE3 - Capacidad para el análisis de datos y la comprensión, modelado y resolución de problemas en contextos de aleatoriedad.
A17	CE17 - Capacidad para la construcción, validación y aplicación de un modelo estocástico de un sistema real a partir de los datos observados y el análisis crítico de los resultados obtenidos.
A20	CE20 - Conocimiento de las herramientas informáticas en el campo del análisis de los datos y modelización estadística, y capacidad para seleccionar las más adecuadas para la resolución de problemas.
B2	CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
B3	CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética
B4	CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado
B6	CG1 - Ser capaz de buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo.
B7	CG2 - Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables.
B8	CG3 - Ser capaz de mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo.
B9	CG4 - Capacidad para abordar con éxito todas las etapas de un proyecto de análisis de datos: exploración previa de los datos, preprocesado, análisis, visualización y comunicación de resultados.
B10	CG5 - Ser capaz de trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones.
C1	CT1 - Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida.
C4	CT4 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad.

Resultados de aprendizaje

Resultados de aprendizaje	Competencias / Resultados del título		
Conocer los principales mecanismos que provocan la falta de datos, la censura en los mismos o la existencia de sesgo en dichos datos	A3 A20	B6	C1 C4
Conocer las principales técnicas estadísticas para analizar problemas con datos faltantes	A3 A17 A20	B3 B4 B9	C1
Conocer las principales técnicas estadísticas para analizar datos funcionales	A3 A17 A20	B3 B4 B9	C1
Conocer las principales técnicas estadísticas para analizar datos censurados	A3 A17 A20	B3 B4 B9	C1
Conocer las principales técnicas estadísticas para analizar problemas con datos sesgados	A3 A17 A20	B3 B4 B9	C1
Ser capaz de aplicar las principales técnicas para datos faltantes, funcionales, censurados y sesgados a conjuntos de datos reales o simulados	A20	B2 B3 B4 B9	C1
Ser capaz de interpretar los resultados y conocer las limitaciones de los métodos	A3	B6 B7 B8 B10	C1 C4

Contenidos	
Tema	Subtema
Introducción al problema de datos faltantes	Retos y problemas ante la falta de datos Mecanismos de falta de datos: missing at random (MAR) y missing completely at random (MCAR) Consecuencias del descarte de los datos faltantes
Técnicas de imputación	Imputación mediante la media Métodos de imputación simple Imputación basada en verosimilitud bajo MAR Algoritmo de Esperanza-Maximización (EM) Métodos de imputación múltiple bajo MAR
Introducción a los datos funcionales	Ejemplos y motivación El registro y la suavización de datos funcionales Métricas y semimétricas para datos funcionales Expresión de los datos funcionales en términos de una base
Análisis de datos funcionales	Estimación de la función media y del operador de covarianzas Concepto de profundidad: detección de datos funcionales atípicos Componentes principales funcionales Modelos lineales para datos funcionales
Datos censurados	Información incompleta y censura Consecuencias de ignorar la censura Estimación paramétrica con datos censurados Estimación no paramétrica: el estimador de Kaplan-Meier El modelo de Cox para la supervivencia condicional

Datos sesgados	Sesgo en la selección de los datos: sesgo por longitud, por tiempo y por tamaño Consecuencias de ignorar el sesgo Estimación de la media y la varianza para datos sesgados El principio de verosimilitud para datos sesgados Situaciones con función de sesgo no especificada

Planificación				
Metodologías / pruebas	Competencias / Resultados	Horas lectivas (presenciales y virtuales)	Horas trabajo autónomo	Horas totales
Presentación oral	A3 B2 B3 B4 C4	21	31.5	52.5
Prácticas a través de TIC	A17 A20 A3 B2 B3 B4 B6 B7 B8 B9 B10 C1	7	24.5	31.5
Trabajos tutelados	A17 A20 A3 B2 B3 B4 B6 B7 B9 B10 C1	3.5	15.75	19.25
Solución de problemas	A17 B2 B7 B8 B10	7	28	35
Prueba mixta	A20 A3 B2 B3 B4 B8 C1	1.5	3	4.5
Prueba mixta	A20 A3 B2 B3 B4 B8 C1	1.5	3.75	5.25
Atención personalizada		2	0	2
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	Descripción
Presentación oral	Presentación con ordenador
Prácticas a través de TIC	Análisis estadístico de conjuntos de datos usando R
Trabajos tutelados	Análisis estadísticos de bases de datos en los que se tengan que aplicar los conceptos estudiados
Solución de problemas	Elección de las herramientas estadísticas y estrategias para resolver problemas con datos faltantes, datos funcionales, datos censurados o datos sesgados
Prueba mixta	Prueba sobre conceptos teóricos y/o ejercicios prácticos con R (a realizar en la mitad del cuatrimestre)
Prueba mixta	Prueba sobre conceptos teóricos y/o ejercicios prácticos con R (a realizar el día del examen oficial)

Atención personalizada	
Metodologías	Descripción
Solución de problemas Prácticas a través de TIC Trabajos tutelados	Asistencia y participación en las clases teóricas Casos prácticos utilizando R Trabajos de análisis de datos Examen sobre conceptos teóricos y/o prácticos

Evaluación			
Metodologías	Competencias / Resultados	Descripción	**Calificación**
Prueba mixta	A20 A3 B2 B3 B4 B8 C1	Prueba de comprensión teórica y aplicacion práctica de los conceptos impartidos (a realizar el día del examen oficial)	40
Trabajos tutelados	A17 A20 A3 B2 B3 B4 B6 B7 B9 B10 C1	Contenido y presentación del trabajo en parejas relacionado con los temas 3, 4 y 5.	30

Prueba mixta	A20 A3 B2 B3 B4 B8 C1	Prueba de comprensión teórica y aplicacion práctica de los conceptos impartidos (a realizar en la mitad del cuatrimestre)	30

Observaciones evaluación
Las puntuaciones de cada parte de la evaluación quedarán de la siguiente forma: Trabajo práctico en parejas relativo a los temas 3-4: 1.75 puntos (1 punto resolución del ejercicio práctico en R y 0.75 presentación oral).Trabajo práctico en parejas relativo al tema 5: 1.25 puntos (0.75 puntos resolución del ejercicio práctico en R y 0.5 presentación oral).Examen de conceptos teóricos/prácticos de los temas 3, 4 y 5: 3 puntos. Tendrá lugar en la mitad del cuatrimestre. Se permite liberar materia, de forma que los estudiantes que obtengan, como mínimo, un 3.5 sobre 10 en este examen parcial, ya no tendrán que examinarse de esta prueba en el examen oficial, al menos que quieran subir nota. Sin embargo, los estudiantes que obtengan una calificación menor a 3.5 sobre 10 o no se presenten al parcial, irán al examen oficial también con esta parte. En el caso de presentarse a subir nota, la calificación que se consideraría relativa a esta prueba sería la obtenida en el examen oficial. Examen de conceptos teóricos/prácticos de los temas 1, 2 y 6: 4 puntos. Tendrá lugar en enero, el día de la convocatoria oficial. Para aprobar la materia, se pide obtener, como mínimo, un 3.5 sobre 10 en esta parte. Para superar la materia será necesario obtener una calificación de, por lo menos, 5 sobre 10 en el conjunto de la materia. En la 2ª oportunidad (julio) los estudiantes deberán hacer las pruebas correspondientes en las que su calificación en la oportunidad de enero hubiera sido inferior a 3.5 sobre 10. En el caso de presentarse a subir nota, la calificación que se consideraría relativa a esta prueba sería la obtenida en el examen oficial de julio. En la primera oportunidad (enero), solo los estudiantes que no se hayan presentado a ninguna de las pruebas evaluables obtendrán la calificación de NO PRESENTADO. En julio obtendrán la calificación de NO PRESENTADO los estudiantes que no se hayan presentado al examen final en esa fecha. Si algún estudiante quiere hacer alguna de las pruebas en un idioma oficial específico (gallego o español), debe avisar al profesorado por lo menos 1 semana antes de la correspondiente prueba.

Fuentes de información	
Básica	- Little R. J., Rubin D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons - Ramsay J. O., Silverman B. W. (2005). Functional Data Analysis. 2nd Edition. Springer - Ferraty F., Vieu P. (2006). Nonparametric functional data analysis : theory and practice. Springer - Hosmer D. W., Lemeshow S., May S. (2008). Applied survival analysis: regression modeling of time-to-event data. Wiley-Interscience - Lee E. T., Wang J. W. (2013). Statistical Methods for Survival Data Analysis. 4th Edition. Wiley - Qin J. (2017). Biased sampling, over-identified parameter problems and beyond (Vol. 5). Springer - Cox D. R. (2005). Some sampling problems in technology. . Selected Statistical Papers of Sir David Cox
Complementária	- Van Buuren, S. (2018). Flexible imputation of missing data. CRC Press - Febrero-Bande M, Oviedo de la Fuente M. (2012). Statistical Computing in Functional Data Analysis: The R Package fda.usc. Journal of Statistical Software, 51(4), 1?28 - Therneau T. M., Grambsch P. M. (2000). Modeling Survival Data: Extending the Cox Model. Springer - Therneau T. (2021). A Package for Survival Analysis in R. CRAN

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Análisis Estadístico de Datos con Dependencia/614G02022 Modelos de Regresión/614G02012 Modelización Estadística de Datos de Alta Dimensión/614G02013 Inferencia Estadística/614G02007 Probabilidad y Estadística Básica/614G02003
Asignaturas que se recomienda cursar simultáneamente
Representación y Gestión de Datos Espacio-Temporales/614G02035 Técnicas de Simulación y Remuestreo/614G02036
Asignaturas que continúan el temario

Gestión de Datos Ómicos y Modelización/614G02042
Otros comentarios

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías