UNIVERSIDADE DA CORUÑA

Guia docente					
Datos Identificativos					2023/24
Asignatura (*)	Modelos de Regresión			Código	614G02012
Titulación	Grao en Ciencia e Enxeñaría de Datos				
Descriptores					
Ciclo	Periodo	Curso		Tipo	Créditos
Grado	1º cuatrimestre	Segundo		Obligatoria	6
Idioma	Castellano				
Modalidad docente	Presencial				
Prerrequisitos					
Departamento	Matemáticas				
Coordinador/a	Vilar Fernandez, Jose Antonio		Correo electrónico	jose.vilarf@udc.es	
Profesorado	Vilar Fernandez, Jose Antonio		Correo electrónico	jose.vilarf@udc.es	
Web					
Descripción general	Estudio de modelos de regresión paramétricos y no paramétricos. En el contexto paramétrico, se abordan procedimientos de inferencia, diagnosis y predicción para el modelo lineal simple, el modelo lineal múltiple y el modelo logístico. La modelización no paramétrica abarcará regresión tipo kernel y por splines, además de una introducción a modelos de regresión parcialmente lineales y a modelos aditivos.				

Competencias / Resultados del título	
Código	Competencias / Resultados del título
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.
A18	CE18 - Capacidad para el análisis de un conjunto de datos y la selección y aplicación de las técnicas de inferencia estadística y de regresión más adecuadas para la adquisición de conocimiento para la toma de decisiones.
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
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.

Resultados de aprendizaje	
Resultados de aprendizaje	Competencias / Resultados del título

Conocer los conceptos generales de la regresión.		B2 B3 B7 B8 B9 B10	
Conocer técnicas de regresión no paramétricas.	A17 A18 A20	B2 B8 B9	C1
Entender la importancia de llevar a cabo una diagnosis de un modelo de regresión.	A17 A18 A20	B2 B3 B8 B9	C1
Saber estimar los parámetros de modelos de regresión lineal simple, múltiple y logístico	A17 A18 A20		C1
Ser capaz de interpretar los resultados y hacer predicciones utilizando modelos de regresión.	A17 A18	B2 B3 B7 B8 B9 B10	
Ser capaz de aplicar las principales técnicas de regresión a conjuntos de datos reales o simulados.	A17 A20	B2 B3 B8 B9	C1
Saber manejar con soltura programas informáticos avanzados de análisis estadístico.	A20	B2 B3 B7 B8 B9 B10	C1

Contenidos	
Tema	**Subtema**
1. Regresión Lineal Simple (RLS). Metodología e Inferencia	1.1 Conceptos generales de regresión 1.2 El modelo RLS. Estimación de los parámetros. Propiedades 1.3 El coeficiente de correlación
2. Regresión Lineal Simple. Diagnosis y Predicción.	2.1 Diagnosis del modelo RLS. Análisis de residuos: linealidad, homocedasticidad, normalidad e independencia 2.2 Observaciones atípicas e influyentes en el modelo RLS 2.3 Transformaciones para conseguir linealidad 2.4 Predicción con el modelo RLS
3. Regresión Lineal Múltiple (RLM). Metodología e Inferencia	3.1 El modelo RLM. Hipótesis básicas del modelo 3.2 Estimación de los parámetros. Propiedades de los estimadores 3.3 Tabla ANOVA. Contraste de la F 3.4 Correlación en regresión múltiple

4. Regresión Lineal Múltiple. Diagnosis y Predicción	4.1 El problema de la multicolinealidad. Definición, efectos, identificación y tratamiento 4.2 Diagnosis del modelo RLM. Análisis de residuos: error de especificación, homocedasticidad, normalidad e independencia 4.3 Robustez del modelo. Observaciones atípicas e influyentes en el modelo RLM 4.4 Predicción con el modelo RLM 4.5 Selección del modelo de regresión. Regresión paso a paso 4.6 Regresión polinómica 4.7 Regresión con variables cualitativas. Variables ficticias
5. Modelo de Regresión con respuesta cualitativa	5.1 Formulación de modelos con respuesta cualitativa 5.2 El modelo logístico 5.3 Estimación y contrastes en el modelo logístico 5.4 Diagnosis del modelo
6. Regresión no paramétrica	6.1 Introducción a los métodos de suavización en regresión 6.2 Estimación núcleo, el estimador de Nadaraya-Watson 6.3 El estimador polinómico local 6.4 Selección del parámetro de suavizado 6.5 Estimadores de tipo spline 6.6 Extensiones. El modelo semiparamétrico (modelo parcialmente lineal). El modelo aditivo

Planificación				
Metodologías / pruebas	**Competencias / Resultados**	**Horas lectivas (presenciales y virtuales)**	**Horas trabajo autónomo**	**Horas totales**
Sesión magistral	A17 A18 B3 B8 B9 B10	30	30	60
Prácticas a través de TIC	A17 A18 A20 B2 B3 B8 B9 C1	20	20	40
Seminario	A18 A20 B2 B3 B8 C1	10	10	20
Solución de problemas	A17 A18 A20 B2 B7 B9 C1	0	20	20
Prueba objetiva	A17 A18 A20 B2 B9 C1	6	0	6
Atención personalizada		4	0	4
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	**Descripción**
Sesión magistral	Se introducirán los fundamentos teóricos de la materia y las principales técnicas para su aplicación en la práctica. Se pondrá particular atención a las condiciones estructurales requeridas para su aplicación y a la diagnosis y evaluación de los modelos ajustados. Su utilidad se ilustrará con ejemplos específicos de diferentes áreas de conocimiento, enfatizando sus ventajas y limitaciones. Se promoverá la participación de los estudiantes.
Prácticas a través de TIC	Se presentarán problemas específicos y los procedimientos de resolución poniendo en práctica conceptos y algoritmos expuestos en las sesiones magistrales. Se tratará de sesiones interactivas donde los problemas propuestos se resolverán con ayuda de scripts con código libre del software R. El docente apoyará y supervisará la puesta en práctica de los conocimientos adquiridos por los estudiantes.
Seminario	Se propondrán cuestiones y ejercicios para discusión y realización conjuntamente con los estudiantes.
Solución de problemas	Se propondrán ejercicios para la realización por parte de los estudiantes de forma individual.

Prueba objetiva	Pruebas objetivas de evaluación de conocimientos.

Atención personalizada	
Metodologías	**Descripción**
Seminario Prácticas a través de TIC	Las prácticas a través de TIC están pensadas para resolver ejercicios usando código en R. Los estudiantes deben familiarizarse con el código y las librerías de R que se usen, comprender y saber interpretar las salidas que se generen y ser capaces de emplearlo individualmente en la ralización de ejercicios similares. Para alcanzar estos objetivos lo más deprisa posible es deseable proporcionar atención personalizada, lo que se realizará durante el desarrollo de las sesiones presenciales y en tutorías individualizadas.

Evaluación			
Metodologías	**Competencias / Resultados**	**Descripción**	**Calificación**
Prueba objetiva	A17 A18 A20 B2 B9 C1	Existirán varias pruebas escritas de conocimientos que constarán de cuestiones teórico-prácticas sobre los contenidos de la materia: -- Pruebas parciales que tendrán lugar una vez que se termine la docencia de bloques temáticos relevantes. Los resultados de estas pruebas formarán parte de la evaluación continua. -- Una prueba final global.	85
Solución de problemas	A17 A18 A20 B2 B7 B9 C1	La participación activa en las clases y la resolución y entrega de ejercicios propuestos a lo largo del curso formará parte de la evaluación continua. Computará hasta el 15% de la calificación.	15

Observaciones evaluación
Presentación y evaluación: 1. Se considerará que un estudiante concurre a una convocatoria cuando participa en actividades que le permiten obtener cuando menos un 50% de la evaluación final.2. Las pruebas parciales de conocimientos tienen carácter voluntario. Si se aprueban, las calificaciones alcanzadas se conservarn, eximiendo al estudiante de contestar las cuestiones relativas a los bloques temáticos ya superados en la prueba final. 3. La prueba final de conocimientos tendrá carácter obligatorio. 4. Será necesario aprobar cada bloque temático para superar la materia. 5. Las calificaciones obtenidas se conservarán entre las oportunidades (ordinaria y extraordinaria) dentro de la convocatoria de cada curso.Segunda oportunidad (prueba de junio/julio) Las calificaciones alcanzadas en las actividades de evaluación continua se guardan para la evaluación en la segunda oportunidad, en la que se realizará un examen global siguiendo las pautas del examen final de primera oportunidad. La nota final en la segunda oportunidad será el máximo de tres cantidades: la nota da evaluación de la primera oportunidad, la nota del nuevo examen (100%), y la media ponderada del nuevo examen (85%) y la parte de evaluación continua distinta de las pruebas parciales de conocimientos (15%).Realización fraudulenta de las pruebas o actividades de evaluación: Una vez comprobada, implicará directamente la calificación de suspenso en la convocatoria en que se cometa: el/la estudiante será calificado con ?suspenso? (nota numérica 0) en la convocatoria correspondiente del curso académico, tanto si la comisión da falta se produce en la primera oportunidad como en la segunda. Para esto, se procederá a modificar su cualificación en el acta de primera oportunidad, si fuese necesario.

Fuentes de información

Básica	- Peña, D. (2002). Regresión y diseño de experimentos. Alianza Editorial - Vilar Fernández, J.M. (2006). Modelos estadísticos aplicados. Universidade da Coruña, Servizo de Publicacións - Faraway, J.J. (2015). Linear models with R . Chapman and Hall - Faraway, J.J. (2006). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman and Hall - Montgomery, D.C., Peck, E.A. y Vining, G.G. (2012). Introduction to linear regression analysis . Wiley - Ritz, C. y Streibig, J.C. (2008). Nonlinear regression with R. Springer - Wand M.P. y Jones M.C. (1995). Kernel Smoothing. Chapman and Hall/CRC - Hosmer, D.W., Lemeshow, S. y Sturdivant, R.X. (2013). Applied logistic regression . John Wiley & Sons
Complementária	- Bowman A.W. y Azzalini A. (1997). Applied Smoothing Techniques for Data Analysis. Oxford University Press - Fox, J. y Weisberg, S. (2011). An R companion to applied regression. SAGE Publications - Agresti, A. (1996). An introduction to categorical data analysis. Wiley - Fan J. y Gijbels I. (1996). Local polynomial modelling and its applications. Chapman and Hall/CRC - Sheather, S.J. (2009). A modern approach to regression with R. Springer - Venables, W.N. y Ripley, B.D. (2010). Modern applied statistics with S . Springer - Wood, S.N. (2006). Generalized Additive Models: An introduction with R. Chapman and Hall/CRC

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Álgebra Lineal/614G02001 Cálculo Multivariable/614G02006 Inferencia Estadística/614G02007 Probabilidad y Estadística Básica/614G02003
Asignaturas que se recomienda cursar simultáneamente
Modelización Estadística de Datos de Alta Dimensión/614G02013
Asignaturas que continúan el temario
Técnicas de Simulación y Remuestreo/614G02036 Análisis Estadístico de Datos Complejos/614G02031 Optimización Matemática/614G02020 Análisis Estadístico de Datos con Dependencia/614G02022
Otros comentarios
La asistencia regular a las clases es altamente recomendable en orden a seguir el desarrollo del programa y participar en la evaluación continua.En cuanto a la aplicación de la vigente normativa en materia de igualdad, conviene subrayar queSe incorporará la perspectiva de género en esta materia (se usará lenguaje no sexista, se propiciará la intervención en clase de alumnos y alumnas...). Se trabajará para identificar y modificar prejuicios y actitudes sexistas y se influirá en el contorno para modificarlos y fomentar valores de respecto e igualdad.Se intentará detectar situaciones de discriminación por razón de género proponiendo acciones y medidas para corregirlas.

(*) 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