



Guía Docente				
Datos Identificativos				2022/23
Asignatura (*)	Modelos de Regresión		Código	614G02012
Titulación	Grao en Ciencia e Enxeñaría de Datos			
Descriptor				
Ciclo	Período	Curso	Tipo	Créditos
Grao	1º cuatrimestre	Segundo	Obrigatoria	6
Idioma	Castelán			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Matemáticas			
Coordinación	Vilar Fernandez, Jose Antonio		Correo electrónico	jose.vilarf@udc.es
Profesorado	Vilar Fernandez, Jose Antonio		Correo electrónico	jose.vilarf@udc.es
Web				
Descrición xeral	Estudo de modelos de regresión paramétricos e non paramétricos. No contexto paramétrico, abóndanse procedementos de inferencia, diagnose e predición para o modelo lineal simple, o modelo lineal múltiple e o modelo loxístico. A modelización non paramétrica comprenderá a regresión tipo kernel e por splines, ademais dunha introdución aos modelos de regresión parcialmente lineais e os modelos aditivos.			

Competencias do título	
Código	Competencias do título
A17	CE17 - Capacidade para a construción, validación e aplicación dun modelo estocástico dun sistema real a partir dos datos observados e a análise crítica dos resultados obtidos.
A18	CE18 - Capacidade para a análise dun conxunto de datos e a selección e aplicación das técnicas de inferencia estatística e de regresión máis adecuadas para a adquisición de coñecemento para a toma de decisións.
A20	CE20 - Coñecemento das ferramentas informáticas no campo da análise dos datos e modelización estatística, e capacidade para seleccionar as máis adecuadas para a resolución de problemas.
B2	CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo
B3	CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética
B7	CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables.
B8	CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo.
B9	CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados.
B10	CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións.
C1	CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.

Resultados da aprendizaxe	
Resultados de aprendizaxe	Competencias do título



Coñecer os conceptos xerais da regresión		B2 B3 B7 B8 B9 B10	
Coñecer técnicas de regresión no paramétricas.	A17 A18 A20	B2 B8 B9	C1
Entender a importancia de levar a cabo a diagnose dun modelo de regresión.	A17 A18 A20	B2 B3 B8 B9	C1
Saber estimar os parámetros dos modelos de regresión lineal simple, múltiple e loxístico.	A17 A18 A20		C1
Ser capaz de interpretar os resultados e facer predicións utilizando modelos de regresión.	A17 A18	B2 B3 B7 B8 B9 B10	
Ser capaz de aplicar as principais técnicas de regresión a conxuntos de datos reais ou simulados.	A17 A20	B2 B3 B8 B9	C1
Saber manexar con soltura programas informáticos avanzados de análise estatística.	A20	B2 B3 B7 B8 B9 B10	C1

Contidos	
Temas	Subtemas
1. Regresión Lineal Simple (RLS). Metodoloxía e Inferencia	1.1 Conceptos xerais de regresión 1.2 O modelo RLS. Estimación dos parámetros. Propiedades 1.3 O coeficiente de correlación
2. Regresión Lineal Simple. Diagnose e Predición.	2.1 Diagnose do modelo RLS. Análise de residuos: linealidade, homoscedasticidade, normalidade e independencia 2.2 Observacións atípicas e influentes no modelo RLS 2.3 Transformacións para conseguir linealidade 2.4 Predición co modelo RLS
3. Regresión Lineal Múltiple (RLM). Metodoloxía e Inferencia	3.1 O modelo RLM. Hipóteses básicas do modelo 3.2 Estimación dos parámetros. Propiedades dos estimadores 3.3 Táboa ANOVA. Contraste F 3.4 Correlación en regresión múltiple



4. Regresión Lineal Múltiple. Diagnose y Predicción	4.1 O problema da multicolinealidade. Definición, efectos, identificación e tratamento 4.2 Diagnose do modelo RLM. Análise de residuos: erro de especificación, homoscedasticidade, normalidade e independencia 4.3 Robustez do modelo. Observacións atípicas e influentes no 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 resposta cualitativa	5.1 Formulación de modelos con resposta cualitativa 5.2 O modelo loxístico 5.3 Estimación e contrastes 5.4 Diagnose do modelo
6. Regresión non paramétrica	6.1 Introducción aos métodos de suavización en regresión 6.2 Estimación núcleo, o estimador de Nadaraya-Watson 6.3 O estimador polinómico local 6.4 Selección do parámetro de suavizado 6.5 Estimadores de tipo spline 6.6 Extensións. O modelo semiparamétrico (modelo parcialmente lineal). O modelo aditivo.

Planificación				
Metodoloxías / probas	Competencias	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Sesión maxistral	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
Proba obxectiva	A17 A18 A20 B2 B9 C1	6	0	6
Atención personalizada		4	0	4
*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado				

Metodoloxías	
Metodoloxías	Descrición
Sesión maxistral	Introduciranse os fundamentos teóricos da materia e as principais técnicas para a súa aplicación na práctica. Particular atención será posta nas condicións estruturais requiridas para a súa aplicación e á diagnose e avaliación dos modelos axustados. A súa utilidade ilustrase con exemplos específicos de diferentes áreas de coñecemento, enfatizando vantaxes e limitacións. Promoverase a participación dos estudantes.
Prácticas a través de TIC	Se presentarán problemas específicos e os procedementos de resolución, poñendo en práctica conceptos e algoritmos expostos nas sesións maxistrais. Tratarase de sesións interactivas donde os problemas propostos se resolverán coa axuda de scripts con código libre do software R. O docente apoiará e supervisará a posta en práctica dos coñecementos adquiridos polos estudantes.
Seminario	Propoñeranse cuestións e exercicios para discusión e realización conxunta cos estudantes.
Solución de problemas	Propoñeranse exercicios para a realización dos estudantes de xeito individual.



Proba obxectiva	Probas obxectivas de avaliación de coñecementos.
-----------------	--

Atención personalizada

Metodoloxías	Descrición
Seminario Prácticas a través de TIC	As prácticas a través de TIC están pensadas para resolver exercicios usando código en R. Os estudantes deben familiarizarse co código e as librerías de R que se usen, comprender e saber interpretar as saídas que se xeneren e ser capaces de empregalo individualmente na realización de exercicios similares. Para acadar estes obxectivos o máis axiña posible é desexable proporcionar atención persoalizada, o que se realizará no desenvolvemento das sesións presenciais e máis en titorías individualizadas.

Avaliación

Metodoloxías	Competencias	Descrición	Cualificación
Proba obxectiva	A17 A18 A20 B2 B9 C1	Existirán varias probas escritas de coñecementos: -- Probas parciais unha vez que se rematen bloques temáticos relevantes. Consistirán en exames escritos. Os resultados destas probas formarán parte da avaliación continua e computarán ata un 25% da cualificación. -- Unha proba de coñecementos final. Será un exame escrito que constará de cuestións teórico-prácticas sobre os contidos da materia. Será obrigatorio e supoñerá o 60% da cualificación.	85
Solución de problemas	A17 A18 A20 B2 B7 B9 C1	A participación activa no desenvolvemento das clases e a resolución e entrega de exercicios propostos ao longo do curso formará parte da avaliación continua. Computará ata o 15% da cualificación.	15

Observacións avaliación

Presentación e avaliación: 1. Considerarase que un estudante concorre a unha convocatoria cando participa en actividades que lle permiten obter cando menos un 50% da avaliación final. 2. A proba final de coñecementos terá carácter obrigatorio. Constará de dúas partes: unha con cuestións de corte teórico e outra máis aplicada e orientada á resolución de problemas. Será necesario aprobar as dúas partes deste exame para supera a materia. 3. A cualificación obtida conservarase entre as oportunidades (ordinaria e extraordinaria) dentro da convocatoria de cada curso. Segunda oportunidade: O peso da avaliación continua na segunda oportunidade será o mesmo que na avaliación ordinaria, un 40%. Na segunda oportunidade de avaliación realizarase un exame e a nota final será o máximo de tres cantidades: a nota da avaliación da primeira oportunidade, a nota do novo exame (100%), e a media ponderada do novo exame (60%) e a avaliación continua (40%).

Fontes de información

Bibliografía 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
----------------------------	---



Bibliografía complementaria	- 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
------------------------------------	--

Recomendacións

Materias que se recomenda ter cursado previamente

Álgebra Lineal/614G02001

Cálculo Multivariable/614G02006

Inferencia Estatística/614G02007

Probabilidade e Estatística Básica/614G02003

Materias que se recomenda cursar simultaneamente

Modelización Estatística de Datos de Alta Dimensión/614G02013

Materias que continúan o temario

Técnicas de Simulación e Remostraxe/614G02036

Análise Estatística de Datos Complexos/614G02031

Optimización Matemática/614G02020

Análise Estatística de Datos con Dependencia/614G02022

Observacións

A asistencia regular as clases é altamente recomendable en orde a seguir o desenvolvemento do programa e participar na avaliación continua.

(*)A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente de acordo coa normativa vixente que establece o proceso de elaboración de guías