



Teaching Guide

Teaching Guide				
Identifying Data				2021/22
Subject (*)	Regression Models		Code	614G02012
Study programme	Grao en Ciencia e Enxeñaría de Datos			
Descriptors				
Cycle	Period	Year	Type	Credits
Graduate	1st four-month period	Second	Obligatory	6
Language	Spanish			
Teaching method	Face-to-face			
Prerequisites				
Department	Matemáticas			
Coordinador	Vilar Fernandez, Jose Antonio	E-mail	jose.vilarf@udc.es	
Lecturers	Vilar Fernandez, Jose Antonio	E-mail	jose.vilarf@udc.es	
	Vilar Fernandez, Juan Manuel		juan.vilar@udc.es	
Web				
General description	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, ademáis dunha introdución aos modelos de regresión parcialmente lineais e os modelos aditivos.			
Contingency plan	1. Modifications to the contents 2. Methodologies *Teaching methodologies that are maintained *Teaching methodologies that are modified 3. Mechanisms for personalized attention to students 4. Modifications in the evaluation *Evaluation observations: 5. Modifications to the bibliography or webgraphy			

Study programme competences

Code	Study programme competences
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.

Learning outcomes			
Learning outcomes		Study programme competences	
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

Contents	
Topic	Sub-topic
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. Hipótesis 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 Predición	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.

Planning				
Methodologies / tests	Competencies	Ordinary class hours	Student?s personal work hours	Total hours
Guest lecture / keynote speech	A17 A18 B3 B8 B9 B10	30	30	60
ICT practicals	A17 A18 A20 B2 B3 B8 B9 C1	20	20	40
Seminar	A20 A18 B2 B3 B8 C1	10	10	20
Problem solving	A17 A18 A20 B2 B7 B9 C1	0	20	20
Objective test	A17 A18 A20 B2 B9 C1	6	0	6
Personalized attention		4	0	4
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description



Guest lecture / keynote speech	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 ilustrarase con exemplos específicos de diferentes áreas de coñecemento, enfatizando vantaxes e limitacións. Promoverase a participación dos estudantes.
ICT practicals	Se presentarán problemas específicos e os procedementos de resolución, poñendo en práctica conceptos e algoritmos expostos nas sesións maxistras. 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.
Seminar	Propoñeranse cuestións e exercicios para discusión e realización conxunta cos estudantes.
Problem solving	Propoñeranse exercicios para a realización dos estudantes de xeito individual.
Objective test	Probas obxectivas de avaliación de coñecementos.

Personalized attention

Methodologies	Description
Seminar ICT practicals	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.

Assessment

Methodologies	Competencies	Description	Qualification
Objective test	A17 A18 A20 B2 B9 C1	A proba de coñecementos final será un exame escrito que constará de varias cuestións teórico-prácticas sobre os contidos da materia. Será obrigatorio e supondrá o 80% da cualificación.	80
Problem solving	A17 A18 A20 B2 B7 B9 C1	A resolución e entrega de exercicios propostos ao longo do curso formará parte da avaliación continua. Computará ata o 20% da cualificación.	20

Assessment comments

Presentación á avaliación: Considérase que un estudante concorre a unha convocatoria cando participa en actividades que lle permiten obter cando menos un 50% da avaliación final. A cualificación obtida conservarase entre as oportunidades (ordinaria e extraordinaria) dentro da convocatoria de cada curso. Segunda oportunidade (proba de xullo) O peso da avaliación continua na segunda oportunidade (proba de xullo) será o mesmo que na avaliación ordinaria, un 20%. 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 (80%) e a avaliación continua (20%).

Sources of information

Basic	- 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 - Peña, D. (2002). Regresión y diseño de experimentos. Alianza Editorial - 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
-------	---



Complementary	- 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
----------------------	--

Recommendations

Subjects that it is recommended to have taken before

Linear Algebra/614G02001

Multivariable Calculus /614G02006

Statistical Inference/614G02007

Probability and Basic Statistics/614G02003

Subjects that are recommended to be taken simultaneously

Statistical Modeling of High Dimensional Data/614G02013

Subjects that continue the syllabus

Simulation and Resampling Techniques/614G02036

Statistical Analysis of Complex Data/614G02031

Mathematical Optimisation/614G02020

Statistical Analysis of Dependent Data/614G02022

Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.