UNIVERSIDADE DA CORUÑA

Teaching Guide			
Identifying Data			**2022/23**
Subject (*)	Data Science	**Code**	710G04026
Study programme	Grao en Xestión Dixital de Información e Documentación		
Descriptors			

Cycle	Period	Year	Type	Credits
Graduate	2nd four-month period	Second	Basic training	6

Language	SpanishGalician		
Teaching method	Face-to-face		
Prerequisites			
Department	Matemáticas		
Coordinador	López Igrexas, Macías	E-mail	macias.lopez@udc.es
Lecturers	González Rueda, Ángel Manuel	E-mail	angel.manuel.rueda@udc.es
	López Igrexas, Macías		macias.lopez@udc.es
Web	https://estudos.udc.es/gl/study/start/710G04V01		
General description	This subject introduces and describes a series of statistical concepts focused on data science. Specifically, it will start with an introduction to sampling techniques and survey design, going through statistical inference as well as the main techniques of multivariate analysis. Additionally, different computational tools related to R statistical software for report generation will be introduced. A mainly applied approach will be used, trying to present all concepts within the context of information and library sciense.		

Study programme competences	
Code	Study programme competences
A1	CE1 - Know and understand the theoretical and methodological principles of information and documentation management to apply them in their professional activity
A8	CE8 - Master the different methods of representation of data, information and knowledge that ensure efficient recovery
A13	CE13 - Know and master the techniques and regulations for the creation and authentication, meeting, selection, organization, representation, preservation, recovery, access, dissemination and exchange, and evaluation of resources and information services
A20	CE20 - Master the bases to develop research activities using multidisciplinary methods and principles
A21	CE21 - Possess knowledge of statistics and quantitative analysis of information
A22	CE22 - Acquire computational skills and management of new ICT
B1	CB1 - Possess and understand knowledge that provides a basis or opportunity to be original in the development and / or application of ideas, often in a research context
B2	CB2 - Apply the knowledge acquired and their ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study
B3	CB3 - Be able to integrate knowledge and face the complexity of making judgments based on information that, being incomplete or limited, includes reflections on social and ethical responsibilities linked to the application of their knowledge and judgments
B4	CB4 - Know how to communicate their conclusions -and the knowledge and ultimate reasons that sustain them- to specialized and non-specialized audiences in a clear and unambiguous way
B5	CB5 - Possess the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous
B6	CG1 - Capacity for cooperation, teamwork and collaborative learning
B7	CG2 - Capacity for reflection and critical reasoning
B8	CG3 - Capacity for planning, organization and management of resources, information and operations
B9	CG4 - Capacity for analysis, diagnosis and decision making
B10	CG5 - Ability to work in an international and global context
B11	CG6 - Ability to understand the importance, value and function of the Digital Information and Documentation Management in the current ICT society
C1	CT1 - Express correctly, both orally and in writing, in the official languages ??of the autonomous community
C2	CT2 - Use the basic tools of information and communication technologies (ICT) necessary for the exercise of their profession and for learning throughout their lives

C3	CT3 - Develop oneself for the exercise of a citizenship that respects democratic culture, human rights and the gender perspective
C4	CT4 - Understand the importance of the entrepreneurial culture and know the means available to entrepreneurs
C5	CT5 - Acquire skills for life and habits, routines and healthy lifestyles
C6	CT6 - Develop the ability to work in interdisciplinary or transdisciplinary teams, to offer proposals that contribute to a sustainable environmental, economic, political and social development
C7	CT7 - Assess the importance of research, innovation and technological development in the socio-economic and cultural progress of society
C8	CT8 - Have the ability to manage time and resources: develop plans, prioritize activities, identify criticisms, establish deadlines and comply with them

Learning outcomes			
Learning outcomes	Study programme competences		
To know the basic inference techniques and acquisition of skills for the estimation and interpretation of confidence intervals and hypothesis testing of one and two populations.	A8 A13 A21	B1 B8 B9	
To know the main types of sampling and the basic tools for survey design.	A1 A13 A20 A21	B2 B3 B4 B5 B9	
Ability to compare two or more populations from databases of different degrees of complexity.	A1 A21	B1 B2 B3 B4 B5	
Knowledge of the different multivariate data analysis techniques to describe and obtain relevant information from complex databases.	A1 A20 A21	B1 B2 B3 B4 B5	
Ability to use computational tools for multivariate data analysis.	A22	B11	C2 C6 C8
Integrate theoretical and practical statistical knowledge as a way to knowledge and reflective and totalizing thinking.	A1 A13	B2 B3 B4 B5 B6 B7 B10	C4 C7 C8
Capacity of analysis and synthesis applied to the management and organization of information.		B2 B3 B4 B5 B6 B7 B8 B9	C1 C3 C5

Acquisition of decision-making skills based on statistical analysis of complex databases.	A21	B2 B3 B8 B9	C8

Contents

Topic	Sub-topic
The following topics develop the contents established in the file of the Verification Report, being:	Introduction and main statistical concepts related to sampling and survey design. Introduction to statistical inference and point estimation. Confidence intervals. Hypothesis testing. Analysis of variance (ANOVA). Regression models. Other multivariate analysis techniques. Computational tools for the generation of statistical reports.
1. Sampling and surveys: introduction and main concepts.	General concepts of statistical sampling and survey design.
2. Introduction to statistical inference and point estimation.	General concepts. Sampling. Parameter estimation. Properties of estimators. Point estimation: point estimation of the mean, variance and a proportion.
3. Confidence intervals.	Concept of confidence interval. Confidence interval for a mean, for a variance, for a proportion and for the difference of two means.
4. Hypothesis testing.	General concepts. Hypothesis testing for the mean, the proportion and for the difference of two means. Independence contrasts.
5. Analysis of variance (ANOVA).	Graphical ANOVA. ANOVA of one factor. ANOVA of more than one factor.
6. Regression models.	Simple and multiple linear regression model. Other regression models.
7. Other multivariate analysis techniques: principal component analysis, factor analysis, correspondence analysis, multidimensional scaling.	Introduction to the most used multivariate techniques.
8. Computational tools for the generation of statistical reports.	Introduction to different tools of the statistical software R for the generation of reports: Rstudio, Rmarkdown, Graphics with R, Htmlwidgets.

Planning

Methodologies / tests	Competencies	Ordinary class hours	Student?s personal work hours	Total hours
Guest lecture / keynote speech	A1 A8 A20 A21 B1 B3 B7 C4 C7	21	0	21
ICT practicals	A13 A22 B11 C2	12	0	12
Case study	A1 A8 A21 B2 B3 B4 B5 B6 B7 B8 B9 C1 C8	7	7	14
Supervised projects	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	1.02	100.98	102
Objective test	A21 B1 B2	1	0	1
Personalized attention		0	0	0
(*)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	Keynote speech will be given in which the teacher will explain, with the help of appropriate audiovisual media, the main contents of the subject.
ICT practicals	In the practical classes the student will be introduced to the handling of the statistical software R. Computational tools for the resolution of problems will be shown and applied through the statistical analysis of data, either from simulated or real data.

Case study	The statistical techniques taught in the course will be applied to solve exercises and real and simulated case studies in the field of digital information management.
Supervised projects	Students will be proposed to develop a group work (2 to 4 people) consisting of the application of statistical and computational tools shown in class to a particular case study, described by real or simulated data. You can also perform a work consisting of the description of a case study in the field of communications and information sciences, in which the resolution of a real problem is carried out based on the application of statistical techniques. A review study on a specific topic of the subject or the software used may also be carried out. The works can be proposed by the teachers or by the students themselves (the proposals will be taken into account or not always according to the teacher's criteria).
Objective test	It will consist of a multiple-choice test on the contents taught in the course.

Personalized attention	
Methodologies	**Description**
Guest lecture / keynote speech ICT practicals	There will be keynote lectures in which the teacher will explain, with the help of appropriate audiovisual media, the main contents of the subject, promoting the debate in class. In the particular case of students with academic dispensation, you can perform face-to-face and virtual tutorials (email, video conference), which allow the student to satisfactorily follow the subject.

Assessment			
Methodologies	**Competencies**	**Description**	**Qualification**
ICT practicals	A13 A22 B11 C2	The attendance and performance of the student in the practical classes will be evaluated, as well as the delivery of works related to the application of the statistical software R.	20
Objective test	A21 B1 B2	It will consist of 15 to 20 test questions with three possible answers.	40
Supervised projects	B2 B4 B5 B6 B8 B9 B10 C1 C3 C5 C6 C8	These works will be carried out in groups of 2 to 5 people, applying statistics to real or simulated data, reviewing a topic on statistics or data science or even regarding a specific application of statistics related to the field of communication and information scienses.	40

Assessment comments
First chance evaluationThere will be a multiple-choice objective test of 10 to 20 questions that represents 40% of the grade. On the other hand, the continuous assessment will consist of attendance and / or delivery of practices related to learning and application of statistical software R for problem solving in the field of digital information management (20% of the overall grade), in addition to the delivery of one and / or several works of application of statistics for the resolution of case studies in library information sciense (alternatively may be tasks of revision or extension of the subject) which represents 40% of the final grade.

Second chance evaluation

The evaluation will be done following the same procedure as in the first opportunity.

Early exam session

All these remarks are applied to the early exam session.

"No presentado" grade

For any of the two opportunities to pass the subject, the "NO PRESENTADO" grade will be given to the students who did non take the objective final test.

Students with recognition of part-time dedication and/or academic exemption of attendance

In the case of students with recognition of part-time dedication and/or academic exemption of attendance that decides not to attend classes, they will be evaluated in the two opportunities as the rest of the students who are in a similar situation.

Fraud in tests or evaluation activities will directly imply the failure grade "0" in the subject in the corresponding call, thus invalidating any grade obtained in all the evaluation activities for the extraordinary call.

Sources of information

Basic	- Daniel Peña (2002). Análisis de datos multivariantes. S.A. MCGRAW-HILL / INTERAMERICANA DE ESPAÑA
	- Cao, R., Francisco, M., Naya, S., Presedo, M.A., Vázquez, M., Vilar, J.A. y Vilar, J.M. (2001). Introducción a la Estadística y sus aplicaciones. Pirámide
	- Egghe, L. y Rousseau, R. (1990). Introduction to Infometrics. Quantitative Methods in Library, Documentation and Information Science. Amsterdam: Elsevier
	- Everitt, B. y Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R. Springer-Verlag New York
Complementary	- Daniel Zelterman (2015). Applied Multivariate Statistics with R. Springer International Publishing
	- Cástor Guisande, Antonio Vaamonde (2012). Gráficos estadísticos y mapas con R. Díaz de Santos
	- Vélez, R. & García, A. (1993). Principios de Inferencia Estadística. UNED

Recommendations
Subjects that it is recommended to have taken before
Fundamentals of Statistics/710G04040
Subjects that are recommended to be taken simultaneously
Subjects that continue the syllabus
Data Mining/710G04030
Other comments
Para axudar a conseguir unha contorna inmediata sostida e cumprir co obxectivo da acción número 5: ?Docencia e investigación saudable e sustentable ambiental e social? do "Plan de Acción Green Campus Ferrol:A entrega dos traballos documentais que se realicen nesta materia:? Solicitaranse en formato virtual e/ou soporte informático.? Realizarase a través de Moodle, en formato dixital sen necesidade de imprimilos.? En caso de ser necesario realizalos en papel:- Non se empregarán plásticos.- Realizaranse impresións a dobre cara.- Empregarase papel reciclado.- Evitarase a impresión de borradores.? Débese de facer un uso sustentable&nbsp; dos recursos e a prevención de impactos negativos sobre o medio natural.? Traballarase para identificar e modificar prexuízos e actitudes sexistas, e influirase na contorna para modificalos e fomentar valores de respecto e igualdade.? Deberanse detectar situacións de discriminación e propoñeranse accións e medidas para corrixilas.

(*)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.