UNIVERSIDADE DA CORUÑA

Teaching Guide			
Identifying Data			2018/19
Subject (*)	Probability. statistics and elements of biomathematics	Code	614522007
Study programme	Mestrado Universitario en Bioinformática para Ciencias da Saúde		

Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	First	Obligatory	6

Language	SpanishGalicianEnglish		
Teaching method	Face-to-face		
Prerequisites			
Department	Matemáticas		
Coordinador		E-mail	
Lecturers	Cao Abad, Ricardo	E-mail	ricardo.cao@udc.es
Web	dm.udc.es/profesores/ricardo/		
General description	It is intended that students acquire skills in identifying situations in which the theory of probability and statistical inference methods are suitable tools for quantitative analysis of databases generated in the field of bioinformatics. To do this, students will complement their knowledge of basic concepts in probability theory and statistical inference; obtain ease in handling the statistical software R, using a large number of resources and introducing the student to programming in this environment. It is intended, also, that students become familiar with the probabilistic models of discrete time stochastic processes and acquire basic training in resampling techniques (bootstrap) as a tool for implementation and evaluation of different statistical algorithms.		

Study programme competences	
Code	Study programme competences
A5	CE5 - Development of skills in the management of statistical techniques and their application to data sets from the bioinformatics field.
A6	CE6 - Ability to identify software tools and most relevant bioinformatics data sources, and acquire skill in their use
A10	CE10 - Draft a bioinformatics research project, anticipating obstacles and possible alternative strategies to resolve them.
B1	CB6 - Own and understand knowledge that can provide a base or opportunity to be original in the development and/or application of ideas, often in a context of research
B4	CB9 - Students should know how to communicate their findings, knowledge and latest reasons underpinning them to specialized and non-specialized audiences in a clear and unambiguous way
B5	CB10 - Students should possess learning skills that allow them to continue studying in a way that will largely be self-directed or autonomous.
C3	CT3 - Use the basic tools of the information technology and communications (ICT) necessary for the exercise of their profession and lifelong learning
C6	CT6 - To assess critically the knowledge, technology and information available to solve the problems they face to.
C8	CT8 - Rating the importance that has the research, innovation and technological development in the socio-economic and cultural progress of society

Learning outcomes			
Learning outcomes	Study programme competences		
G1 - Capacidade para iniciar a investigación e para participar en proxectos de investigación que poden culminar na elabouración duhna teses de doutoramento.	AJ5 AJ6 AJ10	BJ1 BJ4 BJ5	CJ3 CJ6 CJ8
G2 - Capacidade de aplicación de algoritmos de resolución dos problemas e manexo do software adecuado.	AJ5 AJ6 AJ10	BJ1	CJ3

G3 - Capacidade de traballo en equipo e de xeito autónomo	AJ5 AJ6	BJ1 BJ4 BJ5	CJ3 CJ6 CJ8
G4 - Capacidade de formular problemas en termos estatísticos, e de resolvelos utilizando as técnicas axeitadas.	AJ5 AJ6 AJ10	BJ1	CJ3 CJ6
G6 - Capacidade de identificar e resolver problemas	AJ5 AJ6 AJ10	BJ1 BJ5	CJ3
G10 - Capacidade de integrarse nun equipo multidisciplinar para a análise experimental	AJ5 AJ6 AJ10	BJ1 BJ4 BJ5	CJ3 CJ6 CJ8
G11 - Adquirir destreza para o desenvolvemento de software	AJ5 AJ6	BJ5	CJ3
G12 - Capacidade de análise estatística crítica das mostras, os plantexamentos e resultados	AJ5 AJ10	BJ1 BJ5	CJ6 CJ8
G14 - Representar un problema real mediante un modelizado estatístico axeitado.	AJ5 AJ6 AJ10	BJ1 BJ5	
G15 - Deseñar un plano de observación ou recollida de datos que permita abordar o problema de interese	AJ5 AJ6 AJ10	BJ1 BJ5	CJ3 CJ6
E2 - A adquisición dos coñecementos de estatística e investigación de operacións necesarios para a incorporación en equipos multidisciplinares pertencentes a diferentes sectores profesionais.	AJ5 AJ6 AJ10	BJ1 BJ4 BJ5	CJ3 CJ6 CJ8
E4 - Coñecer as aplicacións dos modelos da estatística e a investigación de operacións.	AJ5 AJ10	BJ1 BJ4 BJ5	CJ6
E5 - Coñecer algoritmos de resolución dos problemas e manexar o software axeitado.	AJ5 AJ6 AJ10	BJ1 BJ5	CJ3 CJ6 CJ8
E12 - Realizar inferencias respecto aos parámetros que aparecen no modelo.	AJ5 AJ6 AJ10	BJ1 BJ4 BJ5	CJ3 CJ6 CJ8
E19 - Tratamento de datos e análise estatística dos resultados obtidos.	AJ5 AJ6 AJ10	BJ1 BJ4 BJ5	CJ3
E27 - Obter os coñecementos precisos para unha análise crítica e rigurosa dos resultados.	AJ5 AJ10	BJ1 BJ4 BJ5	CJ6 CJ8
E28 - Complementar a aprendizaxe dos aspectos metodolóxicos con apoio de software.	AJ6 AJ10	BJ5	CJ3 CJ6 CJ8
E78 - Fomentar a sensibilidade cara os principios do pensamento científico, favorecendo as actitudes asociadas ao desenvovemento dos métodos matemáticos, como: o cuestionamento das ideas intuitivas, a análise crítica das afirmacións, a capacidade de análise e síntese ou a toma de decisións racionais	AJ5 AJ10	BJ1 BJ4 BJ5	CJ6 CJ8
E82 - O estudiante será capaz de comprender a importancia da Inferencia Estatística como ferramenta de obtención de información sobre a población en estudo, a partir do conxunto de datos observados dunha mostra representativa de esta. Para iso deberá recoñecer a diferenza entre estatística paramétrica e non paramétrica.	AJ5 AJ10	BJ1 BJ4 BJ5	CJ6 CJ8

E84 - Ser quen de manexar diverso software (en particular R) e interpretar os resultados que proporcionan nos correspondentes estudos prácticos.	AJ5 AJ6 AJ10	BJ4 BJ5	CJ3
E86 - Soltura no manexo da teoría da probabilidade e as variables aleatorias.	AJ5 AJ10	BJ1 BJ4 BJ5	CJ6

Contents	
Topic	**Sub-topic**
1. Basic concepts of probability and statistics revisited.	a. Probability. Random variables and main discrete and continuous distributions. Multivariate distributions. b. Statistical inference: estimation, hypothesis testing and confidence intervals.
2. R statistical programming language revisited.	a. Introduction to R. First steps. Internal functions. Help in R. Functions, loops, vectors. Statistical functions. Plots. Recursivity. R studio. b. Main probability distributions in R. c. Introduction to simulation in R. d. Descriptive statistics in R. e. Hipothesis testing and confidence intervals with R.
3. Linear statistical models.	a. The simple linear regression model. Basic assumptions. Estimation. Testing. Prediction. Model diagnostics. b. The multivariate linear regression model. Basic assumptions. Estimation. Testing. Prediction. Model diagnostics. c. Basic models in experimental desing. One-way and two-way Analysis of Varianza (ANOVA), with or without interaction. Basic assumptions. Estimation. Testing. Model diagnostics. d. The multiple testing problem. False discovery rate.
4. Introduction to stochastic processes.	a. Simple random walk. b. Poisson process and renewal processes. Birth-death processes. c. Markov processes. Markov Chains.
5. Introduction to resampling methods.	a. The uniform Bootstrap. Computing the bootstrap distribution: exact distribution and aproximated distribution using Monte Carlo. Examples. Aplication of the bootstrap for estimating the precision and the bias of an estimator. b. Variations of the uniform Bootstrap. Parametric Bootstrap, symmetrized Bootstrap and smoothed Bootstrap. Discussion and examples. c. Bootstrap methods to construct confidence intervals: percentile method, percentil-t method, simmetrized percentil-t method. Examples. Simulation studies .

Planning				
Methodologies / tests	**Competencies**	**Ordinary class hours**	**Student?s personal work hours**	**Total hours**
Oral presentation	A5 A6 A10 B1 B4 B5 C8	24	36	60
ICT practicals	A5 A6 A10 B4 B5 C3 C6	18	36	54
Multiple-choice questions	A5 B1 B5 C8	1	9	10
Problem solving	A5 A6 A10 B1 B4 B5 C3 C6 C8	4	16	20
Personalized attention		6	0	6
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description
Oral presentation	Presentation computer.
ICT practicals	Datasets statistical analysis using R.
Multiple-choice questions	Multiple-choice test on concepts.
Problem solving	Deciding statistical tools and strategies for problem solving. Linear model formulation. Design of Experiments. Formulation of resampling plans.

Personalized attention	
Methodologies	Description
ICT practicals Problem solving	Attendance and participation in lectures. Written multiple choice test. Participation in workshops and seminars. Practicals to be performed by the student.

Assessment			
Methodologies	Competencies	Description	**Qualification**
ICT practicals	A5 A6 A10 B4 B5 C3 C6	Computer lab using the open statistical software R.	20
Problem solving	A5 A6 A10 B1 B4 B5 C3 C6 C8	Original work on some of the topics of the course concerning some interesting setup in Bioinformatics.	40
Multiple-choice questions	A5 B1 B5 C8	Comprehension Test	40

Assessment comments
The assessment will be carried out using a test on R labs, an individual student work, as well as a written concept test. The concept test score will be 40% of the total qualification, the test on R labs will correspond to 20% of the global score, while the remaining 40% will correspond to the individual student work, that has to be presented orally. To pass the subject is necessary to obtain a score of at least 5 out of 10 overall. On July opportunity, students could avoid those test with scores of at least 4 out of 10 in January tests. Only students that didn't take any test will be qualified as NON ATTENDANT in the first opportunity (January-February). In July opportunity only students that didn't take the final exam will be qualified as NON ATTENDANT.

Sources of information	
Basic	- Cao Abad, R., Francisco Fernández, M., Naya Fernández, S., Presedo Quindimil, M.A., Vázquez Brage, M (2001). Introducción a la Estadística y sus Aplicaciones. Pirámide - Ewens, W.J. and Grant, G.R. (2005). Statistical Methods in Bioinformatics. Springer - Peña Sánchez de Rivera, D. (2000). Estadística: Modelos y Métodos. Alianza Editorial - Ross, S.M. (1995). Stochastic Processes. Wiley - Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall - Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and their Application. Cambridge University Press
Complementary	

Recommendations
Subjects that it is recommended to have taken before
Subjects that are recommended to be taken simultaneously

Introduction to databases/614522002
Genomics/614522006
Fundamentals of bioinformatics/614522008
Introduction to programming/614522001
Foundations of Artificial Intelligence/614522003
Subjects that continue the syllabus
Data structures and algorithmics for biological sequences/614522013
Advanced processing of biological sequences/614522020
Computational intelligence for high dimensional data/614522024
Master thesis/614522025
Computational intelligence for bioinformatics/614522012
Advanced statistical methods in bioinformatics/614522009
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.