



Guía Docente				
Datos Identificativos				2022/23
Asignatura (*)	Probabilidade. estatística e elementos de biomatemática	Código	614522007	
Titulación	Mestrado Universitario en Bioinformática para Ciencias da Saúde			
Descritores				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	1º cuatrimestre	Primeiro	Obrigatoria	6
Idioma	CastelánGalegoInglés			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Matemáticas			
Coordinación	Cao Abad, Ricardo	Correo electrónico	ricardo.cao@udc.es	
Profesorado	Cao Abad, Ricardo	Correo electrónico	ricardo.cao@udc.es	
Web	<a href="http://dm.udc.es/staff/ricardo_cao/">http://dm.udc.es/staff/ricardo_cao/</a>			
Descrición xeral	<p>Preténdese que os alumnos adquiren competencias na identificación de situacións nas que a teoría de probabilidade e os métodos da inferencia estatística son ferramentas axeitadas para a análise cuantitativa de bases de datos xerados na área de bioinformática. Para iso, tratarase de que os estudantes complementen o seu coñecemento dos conceptos básicos de probabilidade e inferencia estatística, obteñan soltura no manexo do software estatístico R, utilizando un gran número de recursos, e que o alumno se introduza na programación nesta contorna. Tamén preténdese que os alumnos se familiaricen cos modelos probabilísticos de procesos estocásticos en tempo discreto e adquiren unha formación básica en técnicas de remostraxe (Bootstrap) como ferramenta para a posta en marcha e avaliación de diferentes algoritmos estatísticos.</p>			

Competencias / Resultados do título	
Código	Competencias / Resultados do título
A5	CE5 ? Desenvolvemento de habilidades no manexo de técnicas estatísticas e a súa aplicación a conxuntos de datos no campo da Bioinformática
A6	CE6 ? Capacidade para identificar as ferramentas software e fontes de datos de bioinformática máis relevantes, e adquirir destreza no seu uso
A10	CE10 - Elaborar un proxecto de investigación bioinformática, anticipando obstáculos e as posibles estratexias alternativas para soluciónalos.
B1	CB6 ? Posuír e comprender o coñecemento que fornecen unha base ou oportunidade de orixinalidade no desenvolvemento e / ou aplicación de ideas, a miúdo nun contexto de investigación.
B4	CB9 - Que os estudantes saiban comunicar as súas conclusións e os coñecementos e principios subxacentes a públicos especializados e non especializados, de xeito claro e inequívoco
B5	CB10 ? Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun xeito que terá de ser en gran parte auto-orientado ou autónomo.
C3	CT3 - 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
C6	CT6 - Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñibles para resolver os problemas cos que deben enfrontarse
C8	CT8 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade

Resultados da aprendizaxe			
Resultados de aprendizaxe		Competencias / Resultados do título	
G2 - Capacidade de aplicación de algoritmos de resolución dos problemas e manexo do software adecuado.		AP5	BP1
		AP6	
		AP10	CP3



G1 - Capacidade para iniciar a investigación e para participar en proxectos de investigación que poden culminar na elaboración dunha teses de doutoramento.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
G3 - Capacidade de traballo en equipo e de xeito autónomo	AP5 AP6	BP1 BP4 BP5	CP3 CP6 CP8
G4 - Capacidade de formular problemas en termos estatísticos, e de resolvelos utilizando as técnicas axeitadas.	AP5 AP6 AP10	BP1	CP3 CP6
G6 - Capacidade de identificar e resolver problemas	AP5 AP6 AP10	BP1 BP5	CP3
G10 - Capacidade de integrarse nun equipo multidisciplinar para a análise experimental	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
G11 - Adquirir destreza para o desenvolvemento de software	AP5 AP6	BP5	CP3
G12 - Capacidade de análise estatística crítica das mostras, os plantexamentos e resultados	AP5 AP10	BP1 BP5	CP6 CP8
G14 - Representar un problema real mediante un modelizado estatístico axeitado.	AP5 AP6 AP10	BP1 BP5	
G15 - Diseñar un plano de observación ou recollida de datos que permita abordar o problema de interese	AP5 AP6 AP10	BP1 BP5	CP3 CP6
E2 - A adquisición dos coñecementos de estatística e investigación de operacións necesarios para a incorporación en equipos multidisciplinares pertencentes a diferentes sectores profesionais.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
E4 - Coñecer as aplicacións dos modelos da estatística e a investigación de operacións.	AP5 AP10	BP1 BP4 BP5	CP6
E5 - Coñecer algoritmos de resolución dos problemas e manexar o software axeitado.	AP5 AP6 AP10	BP1 BP5	CP3 CP6 CP8
E12 - Realizar inferencias respecto aos parámetros que aparecen no modelo.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
E19 - Tratamento de datos e análise estatística dos resultados obtidos.	AP5 AP6 AP10	BP1 BP4 BP5	CP3
E27 - Obter os coñecementos precisos para unha análise crítica e rigurosa dos resultados.	AP5 AP10	BP1 BP4 BP5	CP6 CP8
E28 - Complementar a aprendizaxe dos aspectos metodolóxicos con apoio de software.	AP6 AP10	BP5	CP3 CP6 CP8
E78 - Fomentar a sensibilidade cara os principios do pensamento científico, favorecendo as actitudes asociadas ao desenvolvemento 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	AP5 AP10	BP1 BP4 BP5	CP6 CP8



E82 - O estudante será capaz de comprender a importancia da Inferencia Estatística como ferramenta de obtención de información sobre a poboación 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.	AP5 AP10	BP1 BP4 BP5	CP6 CP8
E84 - Ser quen de manexar diverso software (en particular R) e interpretar os resultados que proporcionan nos correspondentes estudos prácticos.	AP5 AP6 AP10	BP4 BP5	CP3
E86 - Soltura no manexo da teoría da probabilidade e as variables aleatorias.	AP5 AP10	BP1 BP4 BP5	CP6

Contidos	
Temas	Subtemas
1. Revisión de conceptos básicos de probabilidade e estatística.	a. Probabilidade. Variables aleatorias e distribucións notables discretas e continuas. Distribucións multivariantes. b. Inferencia estatística: estimación, contrastes de hipóteses e intervalos de confianza.
2. Revisión da linguaxe de programación estatística R.	a. Introducción ao R. Primeiros pasos. Funcións internas. Axuda en R. Funcións, bucles, vectores. Funcións estatísticas. Gráficas. Recursividade. R studio. b. Principais distribucións de probabilidade en R. c. Introducción á simulación en R. d. Estatística descritiva en R. e. Contrastes de hipóteses e intervalos de confianza con R.
3. Modelos estatísticos lineais.	a. O modelo de regresión linear simple. Hipóteses básicas. Estimación. Contrastos. Predición. Diagnose do modelo. b. O modelo de regresión linear múltiple. Hipóteses básicas. Estimación. Contrastos. Predición. Diagnose do modelo. c. Modelos básicos do deseño experimental. Análise da Varianza (ANOVA) dunha e dúas vías, sen e con interacción. Hipóteses básicas. Estimación. Contrastos. Diagnose do modelo. d. O problema dos contrastes múltiples. False discovery rate.
4. Introducción aos procesos estocásticos.	a. Paseo aleatorio simple. b. Proceso de Poisson e procesos de renovación. Procesos de nacemento e morte. c. Procesos Markovianos. Cadeas de Markov.
5. Introducción aos métodos de remostraxe.	a. O Bootstrap uniforme. Cálculo da distribución Bootstrap: distribución exacta e distribución aproximada por Monte Carlo. Exemplos. Aplicación do Bootstrap á estimación da precisión e o nesgo dun estimador. b. Modificacións do Bootstrap uniforme. Bootstrap paramétrico, simetrizado e suavizado. Discusión e exemplos. c. Métodos bootstrap para a construción de intervalos de confianza: método percentil, percentil-t, percentil-t simetrizado. Exemplos. Estudos de simulación.

Planificación				
Metodoloxías / probas	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas traballo autónomo	Horas totais
Presentación oral	A5 A6 A10 B1 B4 B5 C8	24	36	60



Prácticas a través de TIC	A5 A6 A10 B4 B5 C3 C6	18	36	54
Proba de resposta múltiple	A5 B1 B5 C8	1	9	10
Solución de problemas	A5 A6 A10 B1 B4 B5 C3 C6 C8	4	16	20
Atención personalizada		6	0	6

\*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

Metodoloxías	
Metodoloxías	Descrición
Presentación oral	Presentación con ordenador
Prácticas a través de TIC	Análise estatística de conxuntos de datos usando R.
Proba de resposta múltiple	Proba de resposta múltiple sobre conceptos.
Solución de problemas	Elección das ferramentas estatísticas e estratexias para resolver problemas. Formulación de modelos lineais. Deseño de experimentos. Formulación de plans de remostraxe.

Atención personalizada	
Metodoloxías	Descrición
Prácticas a través de TIC	Asistencia e participación nas clases teóricas. Exame escrito de múltiple opción.
Solución de problemas	Participación en prácticas e seminarios. Suposto práctico a realizar polo alumno.

Avaliación			
Metodoloxías	Competencias / Resultados	Descrición	Cualificación
Presentación oral	A5 A6 A10 B1 B4 B5 C8	Presentación oral do traballo orixinal mencionado no item &quot;Solución de problemas&quot;.	10
Prácticas a través de TIC	A5 A6 A10 B4 B5 C3 C6	Práctica de ordenador usando o software estatístico libre R.	30
Solución de problemas	A5 A6 A10 B1 B4 B5 C3 C6 C8	Traballo orixinal sobre algún dos temas da materia nun contexto de interés en Bioinformática.	30
Proba de resposta múltiple	A5 B1 B5 C8	Proba de comprensión dos conceptos impartidos.	30

Observacións avaliación



A avaliación realizarase por medio dunha proba sobre prácticas con R, un traballo individual do/da alumno/a, así como unha proba escrita de conceptos. A calificación da proba de conceptos representará o 30% da calificación global, a proba de práctica en R corresponderá ao 30% mentres que o 40% restante corresponderá ao traballo individual, que ten que ser presentado en público polos alumnos. A cuarta parte da puntuación deste traballo individual (10% da cualificación global) corresponde á presentación oral do mesmo.

Para superar a materia será necesario obter unha calificación de alomenos 5 sobre 10 no conxunto da materia.

Na

oportunidade de xullo os alumnos poderán liberarse de facer as probas correspondentes nas que a súa calificación na oportunidade de xaneiro fora de alomenos 4 sobre 10.

Na primeira oportunidade (xaneiro-febreiro), so os alumnos que non se teñan presentado a ningunha das probas avaliáveis que figuran

arriba obterán a calificación de NON PRESENTADO. En xullo obterán a calificación de NON PRESENTADO os alumnos que non tiveran presentado ao exame final desa data.

## Fontes de información

<b>Bibliografía básica</b>	<ul style="list-style-type: none"> <li>- 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</li> <li>- Ewens, W.J. and Grant, G.R. (2005). Statistical Methods in Bioinformatics. Springer</li> <li>- Peña Sánchez de Rivera, D. (2000). Estadística: Modelos y Métodos. Alianza Editorial</li> <li>- Ross, S.M. (1995). Stochastic Processes. Wiley</li> <li>- Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall</li> <li>- Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and their Application. Cambridge University Press</li> </ul>
<b>Bibliografía complementaria</b>	

## Recomendacións

### Materias que se recomenda ter cursado previamente

### Materias que se recomenda cursar simultaneamente

Introdución ás bases de datos/614522002  
 Xenómica/614522006  
 Fundamentos de bioinformática/614522008  
 Introdución á programación/614522001  
 Fundamentos de intelixencia artificial/614522003

### Materias que continúan o temario

Estruturas de datos e algoritmia para secuencias biolóxicas/614522013  
 Procesamento avanzado de secuencias biolóxicas/614522020  
 Intelixencia computacional para datos de alta dimensionalidade/614522024  
 Traballo fin de mestrado/614522025  
 Intelixencia computacional para bioinformática/614522012  
 Métodos estatísticos avanzados en bioinformática/614522009

### Observacións

(\*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 dacordo coa normativa vixente que establece o proceso de elaboración de guías

