



| Guía Docente | | | | |
|-----------------------|---|--------------------|--|----------|
| Datos Identificativos | | | | 2023/24 |
| Asignatura (*) | Análise Estatística de Datos Complexos | Código | 614G02031 | |
| Titulación | Grao en Ciencia e Enxeñaría de Datos | | | |
| Descritores | | | | |
| Ciclo | Período | Curso | Tipo | Créditos |
| Grao | 1º cuatrimestre | Cuarto | Optativa | 6 |
| Idioma | CastelánGalego | | | |
| Modalidade docente | Presencial | | | |
| Prerrequisitos | | | | |
| Departamento | Matemáticas | | | |
| Coordinación | Oviedo de la Fuente, Manuel | Correo electrónico | manuel.oviedo@udc.es | |
| Profesorado | López Cheda, Ana Oviedo de la Fuente, Manuel | Correo electrónico | ana.lopez.cheda@udc.es manuel.oviedo@udc.es | |
| Web | https://dm.udc.es/modes/ | | | |
| Descrición xeral | Esta materia proporciona un primeiro contacto do alumnado coas principais técnicas estatísticas para analizar problemas con datos faltantes, datos funcionais, datos censurados ou datos sesgados. Estudaranse os principais mecanismos que provocan a falta de información e aplicaranse as técnicas presentadas a conxuntos de datos reais ou simulados. Analizaranse as limitacións de cada metodoloxía e realizarase diagnose e interpretación dos resultados en termos do problema proposto. | | | |

| Competencias / Resultados do título | |
|-------------------------------------|--|
| Código | Competencias / Resultados do título |
| A3 | CE3 - Capacidade para a análise de datos e a comprensión, modelado e resolución de problemas en contextos de aleatoriedade. |
| 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. |
| 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 |
| B4 | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado |
| B6 | CG1 - Ser capaz de buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo. |
| 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. |
| C4 | CT4 - 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 | | |
| Coñecer os principais mecanismos que provocan a falta de datos, a censura nos mesmos ou a existencia de sesgo en ditos datos | A3 A20 | B6 | C1 C4 |
| Coñecer as principais técnicas estatísticas para analizar problemas con datos faltantes | A3 A17 A20 | B3 B4 B9 | C1 |
| Coñecer as principais técnicas estatísticas para analizar datos funcionais | A3 A17 A20 | B3 B4 B9 | C1 |
| Coñecer as principais técnicas estatísticas para analizar datos censurados | A3 A17 A20 | B3 B4 B9 | C1 |
| Coñecer as principais técnicas estatísticas para analizar problemas con datos sesgados | A3 A17 A20 | B3 B4 B9 | C1 |
| Ser capaz de aplicar as principais técnicas para datos faltantes, funcionais, censurados e sesgados a conxuntos de datos reais ou simulados | A20 | B2 B3 B4 B9 | C1 |
| Ser capaz de interpretar os resultados e coñecer as limitacións dos métodos | A3 | B6 B7 B8 B10 | C1 C4 |

| Contidos | |
|---|--|
| Temas | Subtemas |
| Introdución ós datos funcionais | Exemplos e motivación O registro e a suavización de datos funcionais Métricas e semimétricas para datos funcionais Expresión dos datos funcionais en termos dunha base |
| Análise de datos funcionais | Estimación da función media e do operador de covarianzas Concepto de profundidade: detección de datos funcionais atípicos Compoñentes principais funcionais Modelos lineais para datos funcionais |
| Introdución ó problema de datos faltantes | Retos e problemas ante a falta de datos Mecanismos de falta de datos: missing at random (MAR) e missing completely at random (MCAR) Consecuencias do descarte de datos faltantes |
| Técnicas de imputación | Imputación mediante a media Métodos de imputación simple Imputación baseada en verosimilitude baixo MAR Algoritmo de Esperanza-Maximización (EM) Métodos de imputación múltiple baixo MAR |
| Datos sesgados | Sesgo na selección dos datos: sesgo por lonxitude, por tempo e por tamaño Consecuencias de ignorar o sesgo Estimación da media e a varianza para datos sesgados O principio de verosimilitude para datos sesgados Situacións con función de sesgo non especificada |



| | |
|------------------|--|
| Datos censurados | Información incompleta e censura Consecuencias de ignorar a censura Estimación paramétrica con datos censurados Estimación non paramétrica: o estimador de Kaplan-Meier O modelo de Cox para a supervivencia condicional |
|------------------|--|

| Planificación | | | | |
|---------------------------|--|---|-------------------------|--------------|
| Metodoloxías / probas | Competencias / Resultados | Horas lectivas (presenciais e virtuais) | Horas traballo autónomo | Horas totais |
| Presentación oral | A3 B2 B3 B4 C4 | 21 | 31.5 | 52.5 |
| Prácticas a través de TIC | A17 A20 A3 B2 B3 B4 B6 B7 B8 B9 B10 C1 | 7 | 24.5 | 31.5 |
| Traballos tutelados | A17 A20 A3 B2 B3 B4 B6 B7 B9 B10 C1 | 3.5 | 15.75 | 19.25 |
| Solución de problemas | A17 B2 B7 B8 B10 | 7 | 28 | 35 |
| Proba mixta | A20 A3 B2 B3 B4 B8 C1 | 1.5 | 3 | 4.5 |
| Proba mixta | A20 A3 B2 B3 B4 B8 C1 | 1.5 | 3.75 | 5.25 |
| Atención personalizada | | 2 | 0 | 2 |

*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 |
| Traballos tutelados | Análises estatísticas de bases de datos nas que se teñan que aplicar os conceptos estudados |
| Solución de problemas | Elección das ferramentas estatísticas e estratexias para resolver problemas con datos faltantes, datos funcionais, datos censurados ou datos sesgados |
| Proba mixta | Proba sobre conceptos teóricos e/ou exercicios prácticos con R (a realizar na metade do cuadrimestre) |
| Proba mixta | Proba sobre conceptos teóricos e/ou exercicios prácticos con R (a realizar o día do exame oficial) |

| Atención personalizada | |
|---------------------------|--|
| Metodoloxías | Descrición |
| Solución de problemas | Asistencia e participación nas clases teóricas Casos prácticos utilizando R |
| Prácticas a través de TIC | Traballos de análise de datos Exame sobre conceptos teóricos e/ou prácticos |
| Traballos tutelados | |

| Avaliación | | | |
|---------------------|-------------------------------------|--|---------------|
| Metodoloxías | Competencias / Resultados | Descrición | Cualificación |
| Proba mixta | A20 A3 B2 B3 B4 B8 C1 | Proba de comprensión teórica e aplicación práctica dos conceptos impartidos (a realizar o día do examen oficial) | 40 |
| Traballos tutelados | A17 A20 A3 B2 B3 B4 B6 B7 B9 B10 C1 | Contido e presentación dos traballos prácticos e traballos de investigación | 40 |



| | | | |
|-------------|--------------------------|--|----|
| Proba mixta | A20 A3 B2 B3 B4 B8 C1 | Proba de comprensión teórica e aplicación práctica dos conceptos impartidos (a realizar na metade do cuadrimestre) | 20 |
|-------------|--------------------------|--|----|

Observacións avaliación

As puntuacións de cada parte da avaliación quedarán da seguinte forma:

Dous ou tres traballos prácticos e de investigación: 4 puntos.

Exame de conceptos teóricos/prácticos dos temas 1-2: 2 puntos. Terá lugar na metade do cuadrimestre. Permítese liberar materia, de forma que os estudantes que obtiveran, como mínimo, un 3.5 sobre 10 neste exame parcial, xa non se terán que presentar a esta proba no exame oficial, agás que queiran subir nota. Sen embargo, os estudantes que obtiveran unha calificación menor que 3.5 ou que non se presenten ao parcial, irán ao exame oficial tamén con esta parte. No caso de presentarse a subir nota, a calificación que se consideraría relativa a esta proba sería a obtida no exame oficial.

Exame de conceptos teóricos/prácticos dos temas 3-6: 4 puntos. Terá lugar en xaneiro, o día da convocatoria oficial. Para aprobar a materia, pídese obter, como mínimo, un 3.5 sobre 10 nesta parte.

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

Na 2ª oportunidade (xullo) os estudantes deberán facer as probas correspondentes nas que a súa calificación na oportunidade de xaneiro fose inferior a 3.5 sobre 10. No caso de presentarse a subir nota, a calificación que se consideraría relativa a esta proba sería a obtida no exame oficial de xullo.

Na primeira oportunidade (xaneiro), só os estudantes que non se teñan presentado a ningunha das probas avaliábeis que figuran arriba obterán a calificación de NON PRESENTADO. En xullo obterán a calificación de NON PRESENTADO os estudantes que non se tiveran presentado ao exame final desa data.

Se algún estudante quere facer algunha das probas nun idioma oficial específico (galego ou español), debe avisar ó profesorado alomenos 1 semana antes da correspondente proba.

A realización fraudulenta das probas ou actividades de avaliación, unha vez comprobada, implicará directamente a cualificación de suspenso na convocatoria en que se cometa: o/a estudante será cualificado con ?suspenso? (nota numérica 0) na convocatoria correspondente do curso académico, tanto se a comisión da falta se produce na primeira oportunidade como na segunda. Para isto, procederase a modificar a súa cualificación na acta de primeira oportunidade, se fose necesario.

Fontes de información

| | |
|------------------------------------|---|
| Bibliografía básica | <ul style="list-style-type: none"> - Little R. J., Rubin D. B. (2019). Statistical analysis with missing data (Vol. 793). John Wiley & Sons - Ramsay J. O., Silverman B. W. (2005). Functional Data Analysis. 2nd Edition. Springer - Ferraty F., Vieu P. (2006). Nonparametric functional data analysis : theory and practice. Springer - Hosmer D. W., Lemeshow S., May S. (2008). Applied survival analysis: regression modeling of time-to-event data. Wiley-Interscience - Lee E. T., Wang J. W. (2013). Statistical Methods for Survival Data Analysis. 4th Edition. Wiley - Qin J. (2017). Biased sampling, over-identified parameter problems and beyond (Vol. 5). Springer - Cox D. R. (2005). Some sampling problems in technology. . Selected Statistical Papers of Sir David Cox |
| Bibliografía complementaria | <ul style="list-style-type: none"> - Van Buuren, S. (2018). Flexible imputation of missing data. CRC Press - Febrero-Bande M, Oviedo de la Fuente M. (2012). Statistical Computing in Functional Data Analysis: The R Package fda.usc. Journal of Statistical Software, 51(4), 1?28 - Therneau T. M., Grambsch P. M. (2000). Modeling Survival Data: Extending the Cox Model. Springer - Therneau T. (2021). A Package for Survival Analysis in R. CRAN |

Recomendacións



Materias que se recomenda ter cursado previamente

Análise Estatística de Datos con Dependencia/614G02022
Modelos de Regresión/614G02012
Modelización Estatística de Datos de Alta Dimensión/614G02013
Inferencia Estatística/614G02007
Probabilidade e Estatística Básica/614G02003

Materias que se recomenda cursar simultaneamente

Representación e Xestión de Datos Espazo-Temporais/614G02035
Técnicas de Simulación e Remostraxe/614G02036

Materias que continúan o temario

Xestión de Datos Ómicos e Modelización/614G02042

Observacións

Segundo se recolle nas distintas normativas de aplicación para a docencia universitaria, deberase incorporar a perspectiva de xénero nesta materia (usarase linguaxe non sexista, utilizarase bibliografía de autores/as de ambos sexos, propiciarse a intervención en clase de alumnos e alumnas, etc.) Traballarse 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 por razón de xénero e proponerse accións e medidas para corrixilas.

(*) 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