

Análisis Exploratorio de Datos

Código: 104853
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503852 Estadística Aplicada	FB	1	1

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

Contacto

Nombre: Rosa Camps Camprubí
Correo electrónico: Rosa.Camps@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: catalán (cat)
Algún grupo íntegramente en inglés: No
Algún grupo íntegramente en catalán: Sí
Algún grupo íntegramente en español: No

Equipo docente

Montserrat Ferre Delgado

Prerequisitos

Conocimientos estándar de bachillerato.

Objetivos y contextualización

El objetivo de este curso es, en primer lugar, el conocimiento de las técnicas descriptivas y exploratorias aplicadas a resumir la información contenida en los conjuntos de datos experimentales.

Se empieza con el estudio de una variable, y seguidamente se inicia el estudio comparativo de dos o más variables.

Se dedican apartados especiales al estudio de la evolución temporales de las variables: series cronológicas y números índice.

Un objetivo paralelo, tanto o más importante, es la interpretación de los resultados y de las gráficas en el contexto de los datos.

Competencias

- Analizar datos mediante la aplicación de métodos y técnicas estadísticas, trabajando con datos de diversas tipologías.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- Resumir y descubrir patrones de comportamiento en la exploración de los datos.
- Seleccionar las fuentes y técnicas de adquisición y gestión de datos adecuadas para su tratamiento estadístico.
- Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
- Utilizar correctamente un amplio espectro del software y lenguajes de programación estadísticos, escogiendo el más apropiado para cada análisis y ser capaz de adaptarlo a nuevas necesidades.
- Utilizar eficazmente la bibliografía y los recursos electrónicos para obtener información.

Resultados de aprendizaje

1. Depurar los datos: datos perdidos, transformación de variables, datos anómalos, selección de casos y otras técnicas previas al análisis estadístico.
2. Describir, con los métodos gráficos y analíticos adecuados, datos de tipo cualitativo en una o más variables.
3. Describir, con los métodos gráficos y analíticos adecuados, datos de tipo cuantitativo en una o más variables.
4. Diseñar modificaciones de sintaxis en los programas para realizar nuevos procesos.
5. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
6. Explorar patrones de comportamiento de datos bivariantes.
7. Explorar patrones de comportamiento de datos univariantes.
8. Identificar y seleccionar las fuentes de información más importantes para el análisis descriptivo de datos de diferentes tipologías: sociales, ambientales, sanitarios, económicos, etc.
9. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
10. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
11. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
12. Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
13. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.
14. Utilizar programas estadísticos específicos para el análisis descriptivo de datos.

Contenido

1. Preliminares

- 1.1. Objetivo del análisis exploratorio de datos o estadística descriptiva.
- 1.2. Tipos de variables y escalas de medida.
- 1.3. Redondeos y notación científica.

2. Resumen de datos estadísticos.

- 2.1. Distribuciones de frecuencias: tablas.
- 2.2. Agrupamiento en intervalos.
- 2.3. Representaciones gráficas.

3. Características numéricas de una variable.

- 3.1. Características de posición central: media, mediana, moda.
- 3.2. Otras características de posición: cuartiles, deciles y percentiles.
- 3.3. Características de dispersión: varianza y desviación típica (muestrales y poblacionales), rango, rango inter-cuartil.
- 3.4. Características de dispersión relativa.
- 3.5. Puntuaciones tipificadas.
- 3.6. Características de forma: simetría y curtosis.

4. Complementos en el estudio de una variable.

- 4.1. Análisis exploratorio: diagrama de caja y otros gráficos.
- 4.2. Transformación de variables.
- 4.3. Otras medias: geométrica, armónica, cuadrática.
- 4.4. La desigualdad de Chebyshev.

5. Comparación de una variable en dos o más grupos: Análisis exploratorio

- 5.1. Situación de muestras independientes.
- 5.2. Situación de muestras pareadas

6. Tabulación y representación de la distribución conjunta de los valores de dos variables categóricas.

- 6.1. Tablas de contingencia (distribuciones de frecuencias conjunta, marginales y condicionadas).
- 6.2. Análisis descriptivo de la dependencia entre dos variables categóricas.

7. Descripción numérica de la distribución conjunta de dos variables estadísticas.

- 7.1. Características marginales y condicionadas.
- 7.2. Curvas de regresión y coeficiente de correlación.
- 7.3. Ajuste lineal y predicción.

8. Introducción a las series temporales.

- 8.1. La descomposición clásica.
- 8.2. Suavización de series: aplicación de filtros.

*A menos que las restricciones impuestas por las autoridades sanitarias obliguen a una priorización o reducción de estos contenidos.

Metodología

El trabajo de aula, teoría y problemas, se complementará con prácticas de ordenador donde se utilizará el paquete estadístico R.

En el Moodle del curso los alumnos encontrarán la planificación de la asignatura, las listas de problemas y las prácticas, así como eventuales cambios de aula, horarios, etc.

Es importante tener en cuenta que el CampusVirtual no es una web estática sino que se irá actualizando a lo largo del curso.

Se procurará introducir en la parte más práctica de la asignatura, el análisis y comparación de datos estadísticos por sexo, comentando, en su caso, en el aula las causas y los mecanismos sociales y culturales que pueden sustentar las desigualdades observadas.

*La metodología docente propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Clases de problemas	8	0,32	1, 2, 3, 6, 7, 8, 9, 10, 11
Clases teóricas	18	0,72	5, 1, 2, 3, 6, 7, 8, 9, 10, 11, 13
Estudiar conceptos de teoría, resolver problemas a mano y con R	84	3,36	5, 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14
Prácticas con ordenador	30	1,2	5, 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14

Evaluación

La nota final de la asignatura F se obtendrá a partir de:

- 1) Las notas de los dos exámenes parciales de teoría y problemas, TP1 y TP2, con pesos respectivos 20% y 25%.
- 2) Las notas de las dos pruebas con ordenador, O1 y O2, con pesos respectivos 20% y 25%.
- 3) La asistencia a las sesiones prácticas con ordenador y entregas que se propongan, PC, con un peso del 10%. Esta parte no es recuperable.

La nota final de la asignatura se obtiene haciendo la media ponderada

$$F = 0,2TP1 + 0,2O1 + 0,25TP2 + 0,25O2 + 0,1PC.$$

Requisito para superar la asignatura con la media anterior, las notas TP1, TP2, O1, y O2 deben ser mayores o iguales a 3,7.

En caso de suspender la asignatura, el alumnado tendrá la posibilidad de presentarse a una prueba de síntesis PS.

Para recuperar sólo TP1, TP2, O1, O2. En este caso, su nota final será

$$F = 0,9PS + 0,1PC.$$

Si un estudiante no se presenta en ningún parcial ni a el aprobado de síntesis, se calificará como "No Evaluable".

"Sin perjuicio de otras medidas disciplinarias que se estimen oportunas, y de acuerdo con la normativa Acadèmica vigente, se calificarán con un cero las irregularidades cometidas por el estudiante que puedan conducir a una variación de la calificación de un acto de evaluación. Por lo tanto, plagiar, copiar o dejar copiar una práctica o cualquier otra actividad de evaluación implicará suspender con un cero y no se podrá recuperar en el mismo curso académico. Si esta actividad tiene una nota mínima asociada, entonces la asignatura quedará suspendida. "

Después de las segundas pruebas parciales otorgarán las matrículas de honor que se consideren claras.

Estas matrículas serán ya definitivas. Si el número máximo de matrículas permitido no se ha alcanzado, se reconsiderará la posibilidad de otorgar más después del examen de recuperación.

*La evaluación propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Entrega de ejercicios realizados amb ordinador	10%	2	0,08	5, 1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14
Primer parcial de teoría i problemas	20%	2	0,08	2, 3, 6, 7, 8, 9, 10, 11
Primera prueba de resolución de problemas con ordinador	20%	2	0,08	5, 1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14
Segunda prueba de resolución de problemas con ordinador	25%	2	0,08	5, 1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14
Segundo parcial de teoría i problemas	25%	2	0,08	5, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14

Bibliografía

Manual de teoría

X. BARDINA, M. FARRÉ, Estadística descriptiva, Manuals, 54 Servei de Publicacions, UAB

Bibliografía:

A.J.B. ANDERSON, Interpreting Data. A first cours in Statistics, Ed Chapman and Hall, 1989.

R Tutorial. An R introduction to statistics. (2016). www.r-tutor.com

E. CASA ARUTA, Problemas de Estadística Descriptiva, Ed. Vicens Vives.

R. JOHNSON, P. KUBY, Estadística elemental: Lo esencial, Ed Thomson, 1999.

B. PY, Statistique Descriptive, Ed Económica, 1988.

M. SPIEGEL, Estadística, Teoría y 875 problemas resueltos, Schaum-McGraw-Hill, 1990.

V. ZAIATS, M.L. CALLE i R. PRESAS, Probabilitat i Estadística. Exercicis I, Eumo Ed, 1998.

Bibliografía complementaria.

G. CALOT, Curso de Estadística Descriptiva. Ed Paraninfo, 1988.

FERNÁNDEZ, J.M. CORDERO, A. C'ORDOBA, Estadística Descriptiva, ed ESIC 1996.

L.C HAMMILTON, Modern Data Analysis, Brooks/Cole Publishing Company, 1990.

P.G. HOEL i R.J. JESSEN, Estadística básica para negocios y economía, Compañía Editorial Continental, Mexico, 1993.

R.K. PEARSON, Exploratory Data Analysis using R. Data Mining and Knowledge Discovery Series, Chapman & Hall/CRC, 2018.

D. PEÑA SÁNCHEZ DE RIVERA, Estadística. Modelos y métodos. 1. Fundamentos i 2. Modelos lineales y series temporales, Alianza Editorial 1995. (2 volúmenes)

Software

R i RStudio