UAB Universitat Autònoma de Barcelona

Análisis de Datos Complejos

2026/2027

Código: 104399
Créditos: 6

Titulación	Tipo	Curso
Matemática Computacional y Analítica de Datos	OB	2

Profesor/a de contacto

Nombre : Amanda Fernandez Fontelo

Correo electrónico : amanda.fernandez@uab.cat

Equipo docente

Alan Morte Piferrer

Idiomas de los grupos

Puede consultar esta información al final del documento.

Prerrequisitos

Se recomienda tener conocimientos de probabilidad e inferencia estadística así como una cierta soltura con el programa R.

Objetivos

El principal objetivo es proporcionar herramientas estadísticas para el análisis de datos, dominando las técnicas más relevantes para enfrentarse con modelos complejos.

Resultados de aprendizaje

- CM14 (Implementar estrategias para confirmar o refutar hipótesis.) Implementar estrategias para confirmar o refutar hipótesis.
- CM15 (Gestionar la información para su validación mediante un tratamiento estadístico.) Gestionar la información para su validación mediante un tratamiento estadístico.
- CM16 (Evaluar, a partir de datos obtenidos, desigualdades por razón de sexo/género.) Evaluar, a partir de datos obtenidos, desigualdades por razón de sexo/género.
- KM12 (Identificar la inferencia estadística como instrumento de pronóstico y predicción.) Identificar la inferencia estadística como instrumento de pronóstico y predicción.
- KM14 (Identificar la utilidad de los métodos Bayesianos, aplicándolos oportunamente.) Identificar la utilidad de los métodos Bayesianos, aplicándolos oportunamente.
- SM14 (Utilizar las propiedades de las funciones de densidad y de distribución.) Utilizar las propiedades de las funciones de densidad y de distribución.
- SM15 (Utilizar software estadístico apropiado para gestionar bases de datos, para obtener índices de resumen de las variables del estudio y para analizar los datos mediante técnicas de inferencia.) Utilizar software estadístico apropiado para gestionar bases de datos, para obtener índices de resumen de las

variables del estudio y para analizar los datos mediante técnicas de inferencia.

Contenidos

Tema 1- Modelos lineales: regresión lineal múltiple y ANOVA.
Tema 2- Modelos lineales generalizados: regresión logística y de Poisson.
Tema 3- Técnicas de Big Data en modelos lineales y modelos lineales generalizados.
Tema 4- Técnicas de remuestreo: Bootstrap.
Tema 5 (si hay tiempo)- Regularización: Modelos lasso y ridge.

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Prácticas de laboratorio (PLAB)	14	0,56	CM14, CM15, CM16, SM15
Sesiones de seguimiento y tutorías	4	0,16	CM14, CM15, CM16, KM12, SM14, SM15
Sesiones de teoría (TE)	27	1,08	CM14, CM15, CM16, KM12, SM14, SM15
Ampliación de conceptos introducidos en las sesiones de teoria	46	1,84	CM14, CM15, CM16, KM12, SM14
Resolución de problemas en clase (SEM)	8	0,32	CM14, CM15, CM16, KM12, SM14, SM15
Solución de problemas teóricos	28	1,12	CM14, CM15, CM16, KM12, SM14
Realización autónoma de cada práctica de laboratorio	16	0,64	CM14, CM15, CM16, KM12, SM14, SM15

De acuerdo con los objetivos de la asignatura, el desarrollo del curso se basa en las siguientes actividades:

- Sesiones de teoría: El alumno adquiere los conocimientos científico-técnicos propios de la asignatura, asistiendo en las clases de teoría, complementándolas con el estudio personal de los temas explicados. Las clases de teoría son las actividades en las cuales se exige menos interactividad al estudiante: están concebidas como un método fundamentalmente unidireccional de transmisión de conocimientos del profesor al alumno. Las clases se harán utilizando un apoyo de diapositivas en inglés que se colgarán también en el Moodle del curso. Se deben complementar estas diapositivas con algún libro de referencia del curso.
- Sesiones de problemas y prácticas: Los problemas y las prácticas son sesiones con una doble misión. Por un lado, se trabajan los conocimientos científico-técnicos expuestos en las clases de teoría para completar su comprensión y profundizar en ellos, desarrollando actividades diversas, desde la típica resolución de problemas hasta la discusión de casos prácticos. Por otro lado, las clases de problemas son el foro natural en el cual discutir en común el desarrollo del trabajo práctico, aportando los conocimientos necesarios para llevarlo adelante, o indicando donde y como se pueden adquirir. Aunque tanto los problemas como las practicas se trabajan en las sesiones presenciales correspondientes, estos también se deben trabajar de manera autónoma en casa. Tanto las sesiones de problemas como las sesiones de prácticas tienen un tiempo limitado que no permite trabajar todos los problemas y casos prácticos necesarios para adquirir satisfactoriamente todos los contenidos del curso. Por lo tanto, el trabajo autónomo tanto de problemas como de prácticas es estrictamente necesario para conducir las correspondientes sesiones de la manera más eficiente posible.

El curso práctico de esta asignatura se plantea como un camino para orientar el estudiante en un trabajo de campo de estadística en cada una de sus etapas. Así realizan por su cuenta con el software R unos ejercicios

prácticos dirigidos a resolver problemas reales concretos. Este planteamiento está orientado a promover un aprendizaje activo y a desarrollar el razonamiento crítico y la capacidad de análisis y síntesis.

En esta asignatura se permite el uso de tecnologías de Inteligencia Artificial (IA) como parte del desarrollo de las actividades, siempre que el resultado final refleje una contribución significativa del estudiante en el análisis, el razonamiento y la reflexión personal. El estudiante deberá identificar claramente qué partes del trabajo han sido generadas o asistidas mediante estas tecnologías, especificar las herramientas empleadas e incluir una breve reflexión crítica sobre la influencia que estas han tenido en el proceso de elaboración y en el resultado final de la actividad. La falta de transparencia en el uso de herramientas de IA se considerará una vulneración de los principios de honestidad académica y podrá conllevar una penalización en la calificación de la actividad o, en los casos de mayor gravedad, la aplicación de las medidas disciplinarias previstas por la normativa de la Universitat Autònoma de Barcelona.

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.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Examen Bootstrap (BE)	15	2	0,08	CM14, CM15, CM16, KM12, KM14, SM14, SM15
Examen 2 (E2)	20	2	0,08	CM14, CM15, KM12, SM14
Problemas y prácticas evaluables en el aula (PP)	30	0	0	CM14, CM15, KM12, SM14, SM15
Examen 1 (E1)	35	3	0,12	CM14, CM15, KM12, SM14

Ver versión en catalán.

Bibliografía

- Introduction to Linear Regression Analysis. Montgomery, D. Peck, A. Vining, G., 2001.
- An R Companion to Linear Statistical Models. Christopher Hay-Jahans, 2012.
- Generalized Linear Models. McCullagh, P. and Nelder, J., 1992.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Hastie T., Tibshirani, R., Friedman, J. 2009.
- Resampling methods: a practical guide to data Analysis. Phillip I. Good, 2006.
- The jackknife, the bootstrap and other resampling plans. Bradley Efron, 1982.
- Bootstrap methods and their application. A.C. Davison, D.V. Hinkley, 1997.

Software

Se utilizará el programa R.

Grupos e idiomas de la asignatura

La información proporcionada es provisional hasta el 30 de noviembre. A partir de esta fecha, podrá consultar el idioma de cada

grupo a través de este enlace. Para acceder a la información, será necesario introducir el CÓDIGO de la asignatura

Tipo de docencia	Grupo	Idioma	Semestre	Turno
(TE) Teoría	1	Catalán/Español	segundo cuatrimestre	manaña-mixto
(PLAB) Prácticas de laboratorio	1	Catalán/Español	segundo cuatrimestre	manaña-mixto
(SEM) Seminarios	1	Catalán/Español	segundo cuatrimestre	manaña-mixto