

Guía docente

240615 - 240615 - Introducción a Data Science

Última modificación: 20/05/2025

Unidad responsable: Escuela Técnica Superior de Ingeniería Industrial de Barcelona
Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa.

Titulación: GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES (Plan 2010). (Asignatura optativa).

Curso: 2025 **Créditos ECTS:** 4.5 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: JOSEP GINEBRA

Otros: JOSEP GINEBRA

CAPACIDADES PREVIAS

Haber aprobado Estadística.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

1. Conocimientos básicos sobre el uso y programación de los ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería.
2. Conocimientos y capacidades para organizar y gestionar proyectos. Conocer la estructura organizativa y las funciones de una oficina de proyectos.

Transversales:

3. COMUNICACIÓN EFICAZ ORAL Y ESCRITA: Comunicarse de forma oral y escrita con otras personas sobre los resultados del aprendizaje, de la elaboración del pensamiento y de la toma de decisiones; participar en debates sobre temas de la propia especialidad.
4. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro más, o realizando tareas de dirección con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

METODOLOGÍAS DOCENTES

Todas las clases son prácticas, delante de la computadora. El análisis de los datos se hará con MINITAB y con R. Cada semana se propondrán pequeños ejercicios de análisis de datos a realizar en casa. Se tendrá que realizar un trabajo fin de asignatura.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

Al terminar la asignatura el estudiante deberá ser capaz de identificar situaciones en las que es provechoso analizar datos, de identificar el modelo o método de análisis mas adecuado para sus datos, de construir un modelo que sintetice la información en los datos y le permita hacer predicciones, de reducir la dimensionalidad y visualizar datos multivariantes, de implementar algoritmos de clasificación supervisada y no supervisada, y de evaluar la cualidad de los resultados obtenidos.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas grupo mediano	45,0	40.00
Horas aprendizaje autónomo	67,5	60.00

Dedicación total: 112.5 h

CONTENIDOS

Tema 1: Introducción

Descripción:

(CAST) 1.- Modelización estadística. 2.- Análisis multivariante.

Dedicación: 3h 30m

Grupo grande/Teoría: 1h 30m

Actividades dirigidas: 1h

Aprendizaje autónomo: 1h

Tema 2: Modelos lineales

Descripción:

(CAST) 1.- Modelo lineal normal. 2.- Ajuste por mínimos cuadrados y otros criterios. 3. Descomposición ANOVA y bondad de ajuste. 4.- Intervalos de confianza y tests sobre los coeficientes. 5.- Predicción. 6.- Validación del modelo via análisis de residuos. 7.- Selección del modelo y validación cruzada. 8.- Uso de variables explicativas categóricas. 9.- Interpretación del modelo; Colinealidad, sesgo y causalidad.

Dedicación: 30h

Grupo grande/Teoría: 6h

Grupo pequeño/Laboratorio: 6h

Actividades dirigidas: 6h

Aprendizaje autónomo: 12h

Tema 3: Modelos no-lineales

Descripción:

1.- Model no-lineal normal teòric i ajustat. 2.- Ajust del model. 3.- Intervals de confiança i tests. 4.- Validació del model.

Dedicación: 6h

Grupo grande/Teoría: 1h 30m

Grupo pequeño/Laboratorio: 1h 30m

Aprendizaje autónomo: 3h

Tema 4: Modelos para respuesta categórica y para conteos

Descripción:

(CAST) 1.- Model per a resposta binària. 2.- Ajust. 3.- Intervals de confiança i tests sobre els coeficients. 4.- Validació. 5.- Selecció de models i validació creuada. 6.- Interpretació del model. 7.- Model per a resposta politòmica. 8.- Taules de contingència i model logístic. 9.- Model per a comptetjos.

Dedicación: 22h 30m

Grupo grande/Teoría: 4h 30m

Grupo pequeño/Laboratorio: 4h 30m

Actividades dirigidas: 4h 30m

Aprendizaje autónomo: 9h

Tema 5: Modelos de series temporales

Descripción:

1.- Descripció d'una sèrie temporal; Estacionarietat i estacionalitat. 2.- Models AR. 3.- Models MA. 4.- Modelos ARIMA. 5.- Modelos ARIMA estacionales.

Dedicación: 13h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 3h

Actividades dirigidas: 3h

Aprendizaje autónomo: 4h

Tema 6: Visualización de datos multivariantes (Reducción de la dimensionalidad)

Descripción:

(CAST) 1.- Anàlisi de components principals. 2.- Anàlisi de correspondències.

Dedicación: 6h

Grupo grande/Teoría: 1h 30m

Grupo pequeño/Laboratorio: 1h 30m

Aprendizaje autónomo: 3h

Tema 7: Anàlisi clúster (Clasificación no supervisada)

Descripción:

1.- Mètodes jeràrquics. 2.- Mètodes de partició (Algoritmes k-means). 3.- Anàlisi cluster per variables.

Dedicación: 6h 30m

Grupo grande/Teoría: 1h 30m

Grupo pequeño/Laboratorio: 1h 30m

Actividades dirigidas: 1h 30m

Aprendizaje autónomo: 2h

Tema 8: Anàlisi discriminant (Clasificación supervisada)

Descripció:

1.- Anàlisi discriminant lineal. 2.- Comparació de mètodes de classificació. 3.- Anàlisi discriminant quadràtic. 4.- Anàlisi discriminant logístic. 5.- Mètodes de classificació en arbre. 6.- Nearest neighbor classifier. 7.- Sensitivitat, especificitat i la corba ROC.

Dedicació: 13h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 3h

Actividades dirigidas: 3h

Aprendizaje autónomo: 4h

SISTEMA DE CALIFICACIÓN

Habrà un examen parcial hecho en casa y un examen final hecho en clase.

Nota Asignat. = 0,1 Ejercicios + 0,3 Proyecto + 0,1 Examen Parcial + 0,5 Examen Final

Durante el cuatrimestre de primavera de 2019-20, y como consecuencia de la crisis del covid 19, el sistema de calificación será el mismo, con la única diferencia de que el examen final sera un take home en vez de hacerlo en clase.

BIBLIOGRAFÍA

Básica:

- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning : data mining, inference and prediction [en línea]. 2nd ed. New York: Springer, cop 2009 [Consulta: 28/04/2025]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7>. ISBN 0387952845.
- Peña, Daniel. Regresión y diseño de experimentos. Madrid: Alianza, 2002. ISBN 9788420693897.
- Venables, William N; Ripley, B.D. Modern Applied Statistics with S. 4th ed. New York: Springer Verlag, 2003. ISBN 0387954570.
- Peña, Daniel. Análisis de datos multivariantes [en línea]. 1a ed. Madrid: McGraw-Hill/Interamericana de España, S.L., 2013 [Consulta: 22/10/2024]. Disponible a: https://www-ingebook-com.recursos.biblioteca.upc.edu/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203. ISBN 9788448191849.
- Weisberg, Sanford. Applied linear regression. 4th ed. Hoboken, New Jersey: Wiley, 2014. ISBN 9781118594858.
- Everitt, B.S.; Dunn, G. Applied multivariate data analysis [en línea]. 2nd. Chichester, West Sussex: Wiley, 2001 [Consulta: 24/10/2024]. Disponible a: <https://onlinelibrary-wiley-com.recursos.biblioteca.upc.edu/doi/epdf/10.1002/9781118887486>. ISBN 9781118887486.
- Greenacre, Michael J. Correspondence analysis in practice [en línea]. 3rd ed. Boca Raton: CRC Press, Taylor & Francis Group, 2017 [Consulta: 20/02/2025]. Disponible a: <https://www-taylorfrancis-com.recursos.biblioteca.upc.edu/books/mono/10.1201/9781315369983/correspondence-analysis-practice-michael-greenacre>. ISBN 9781498731782.
- Gareth James,... [et al.]. An Introduction to statistical learning : with applications in R [en línea]. 2nd ed. New York: Springer, 2021 [Consulta: 01/10/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6686746>. ISBN 1071614177.

Complementaria:

- Clarke, B.; Fokoue, E.; Zhang, H.H. Principles and theory for data mining and machine learning [en línea]. 1st ed. New York: Springer, 2009 [Consulta: 26/08/2022]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-98135-2>. ISBN 9780387981345.
- Dobson, Annette J; Barnett, Adrian G. An Introduction to generalized linear models. 4th ed. Boca Raton: CRC Press / Chapman & Hall, 2018. ISBN 9781138741515.
- Johnson, Richard; Wichern, Dean. Applied multivariate statistical analysis [en línea]. Pearson new international edition. Harlow, Essex: Pearson, 2014 [Consulta: 24/10/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5174>

[865](#). ISBN 9781292024943.

- Peña, Daniel. Análisis de series temporales. 2nd ed. Madrid: Alianza, cop. 2010. ISBN 9788420669458.

- Wakefield, Jon. Bayesian and frequentist regression methods [en línea]. 1st ed. New York, NY: Springer, 2013 [Consulta: 11/07/2025]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-1-4419-0925-1>. ISBN 9781441909244.