

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	36502
Nom	Anàlisi Exploratòria de Dades i Base de Dades
Cicle	Grau
Crèdits ECTS	6.0
Curs acadèmic	2022 - 2023

Titulació/titulacions

Titulació	Centre	Curs	Període
1332 - Grau en Intel·ligència i Analítica de Negocis/BIA	Facultat d'Economia	1	Primer quadrimestre

Matèries

Titulació	Matèria	Caràcter
1332 - Grau en Intel·ligència i Analítica de Negocis/BIA	5 - Fonaments de l'Anàlisi de Dades	Formació Bàsica

Coordinació

Nom	Departament
COLL SERRANO, VICENTE	110 - Economia Aplicada

RESUM

Anàlisi exploratòria de dades i bases de dades és una assignatura de formació bàsica adscrita a l'àrea de *Mètodes quantitius per a l'economia i l'empresa*, que s'imparteix en el primer semestre del primer curs del grau d'intel·ligència i analítica de negocis, amb una càrrega lectiva total de sis crèdits ECTS.

L'anàlisi exploratòria de dades (AED), que va crear John W. Tukey, és la primera etapa del procés d'anàlisi de dades. Consisteix en un conjunt de mètodes estadístics i gràfics que permeten:

- Organitzar i estructurar les dades.
- Explorar la distribució de les variables considerades amb la finalitat, per exemple, de detectar valors perduts i valors anòmals, comprendre la dispersió i la forma de les dades, etc.
- Entendre les relacions entre les variables.
- Comprovar el compliment dels supòsits en què es basen gran part dels mètodes multivariants.
- Reorganitzar i reestructurar les dades per a posteriors procediments d'anàlisi.



Per capacitar l'estudiant en l'anàlisi exploratòria de dades s'estudien els conceptes bàsics d'estadística descriptiva de caràcter tant unidimensional com multidimensional, i es promou el treball amb bases de dades reals utilitzant com a "sorftware" d'anàlisi estadística el programa R i RStudio.

CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

No es requereixen.

COMPETÈNCIES

1332 - Grau en Intel·ligència i Analítica de Negocis/BIA

- Que els estudiants tinguin la capacitat d'arreglar i interpretar dades rellevants (normalment dins de la seua àrea d'estudi) per emetre judicis que incloguen una reflexió sobre temes rellevants d'índole social, científica o ètica.
- Capacitat d'accés i gestió de la informació en diferents formats per a la seva posterior anàlisi a fi d'obtenir coneixement a través de dades.
- Capacitat d'anàlisi i síntesi.
- Capacitat per a analitzar i buscar informació provinent de fonts diverses.
- Capacitat per a utilitzar les TIC, tant en l'àmbit d'estudi com en el desenvolupament professional.
- Aplicar mètodes i tècniques d'anàlisis, síntesis i representació gràfica mitjançant programes informàtics.
- Conèixer les diferents tipologies de dades.
- Reorganitzar i reestructurar variables i bases de dades.

RESULTATS DE L'APRENTATGE

- Conèixer les diferents tipologies de dades amb què es pot treballar per resoldre problemes reals i identificar-les.
- Saber accedir a dades de diferent tipologia.
- Aplicar diferents mètodes i tècniques d'anàlisi exploratòria mitjançant l'ús de programes informàtics (R i RStudio).
- Conèixer el raonament lògic/estratègic per abordar situacions reals del món econòmic.
- Saber buscar, organitzar, seleccionar i valorar informació.
- Saber organitzar i depurar dades i bases de dades i com fusionar-les.
- Manejar eines quantitatives bàsiques i la seua aplicació a l'entorn econòmic.
- Augment de la capacitat per a identificar, classificar, raonar, argumentar i interpretar les relacions



entre les variables econòmiques.

- Augment de l'habilitat per a utilitzar el raonament lògic/estratègic per a abordar situacions reals del món econòmic.

DESCRIPCIÓ DE CONTINGUTS

1. Tema 1. Introducció a l'Anàlisi Exploratori de Dades

2. Tema 2. R per a l'anàlisi estadística de dades. Introducció.

- Tipus de dades en R.
- Vectors, matrius, llistes i dataframes.

3. Tema 3. Accés, neteja i depuració de dades.

- Fitxers de text: csv, txt.
- Formats propietaris: Excel, SPSS, Stata, etc.
- Formats de R: RData, rds.
- Accés a dades a través de API, dades no estructurades, etc.
- Neteja de dades: el procés de neteja i depuració de dades

4. Tema 4. Iniciació a la visualització de dades.

- Histograma.
- Diagrama de barres.
- Diagrama de caixa (boxplot).
- Gràfics de línies.

5. Tema 5. Transformació de dades.

- Què són les dades ordenades?
- Reorganització de les dades: com passar de format ample a format llarg i al contrari.
 - o `gather()`
 - o `spread()`
 - o `separate()`
 - o `unite()`

**6. Tema 6. Anàlisi de dades multidimensionals.**

- Taules de freqüències conjuntes i marginals.
- Vector de mesures centrals i matriu de variàncies-covariàncies.
- Relació entre variables (covariància, correlació, etc.)
- Regressió i predicció.
- Valors absents i imputació.

7. Tema 7. Data wrangling: maneig/gestió de les dades.

- Gramàtica per al maneig de dades amb R: el paquet dplyr.
 - o select()
 - o filter()
 - o group_by()
 - o summarize()
 - o arrange()
 - o mutate()
- Unió/fusió de bases de dades:
 - o Ideal: r_bind() i c_bind()
 - o Mutating joins: inner_join(), left_join(), right_join(), full_join()

VOLUM DE TREBALL

ACTIVITAT	Hores	% Presencial
Pràctiques en aula informàtica	45,00	100
Classes de teoria	15,00	100
Elaboració de treballs individuals	10,00	0
Estudi i treball autònom	26,00	0
Preparació d'activitats d'avaluació	14,00	0
Preparació de classes de teoria	10,00	0
Preparació de classes pràctiques i de problemes	20,00	0
Resolució de casos pràctics	10,00	0
TOTAL	150,00	

METODOLOGIA DOCENT

El desenvolupament de l'assignatura s'estructura fonamentalment en sessions teòriques i sessions pràctiques. Segons el tipus de sessió (teòrica o pràctica) s'elegirà un mètode didàctic o altre.



En les sessions teòriques, d'una durada d'una hora, s'exposaran els continguts teoricopràctics corresponents als fonaments de l'estadística descriptiva i de l'anàlisi exploratòria de dades.

El mètode docent predominant en les classes teòriques serà la classe magistral participativa. Aquesta metodologia permet dirigir de forma organitzada els grups grans d'estudiants i oferir els avantatges d'una classe magistral sense limitar la participació dels estudiants i la relació professor-estudiant. S'intentarà fomentar la participació i la discussió en la classe amb la finalitat d'oferir a l'alumne un contacte directe amb el contingut.

En les sessions pràctiques, que tenen una durada de tres hores, el professor guiarà els estudiants en el maneig dels programes R i RStudio i la seua aplicació a l'anàlisi exploratòria de dades a través de l'ús de bases de dades reals. Per afermar el coneixement adquirit en l'assignatura es proposarà als estudiants la resolució de casos pràctics similars, però no idèntics als que es tracten en classe. Per la naturalesa de les sessions pràctiques és indispensable el seguiment setmanal dels continguts estudiats per part dels estudiants.

Les sessions pràctiques es desenvoluparan amb diferents estratègies docents en funció dels continguts de les sessions teòriques i pràctiques.

AVALUACIÓ

L'assignatura s'avaluarà per avaluació contínua i amb un examen de síntesi al final del quadrimestre.

L'examen de síntesi, no necessàriament de caràcter escrit, constarà de preguntes teoricopràctiques que permetran valorar si l'estudiant ha assimilat els conceptes clau del programa. Aquest examen de síntesi representa el 70% de la nota final.

L'avaluació contínua té l'objectiu de desenvolupar les competències dels estudiants i estimular el treball diari. Per tant, aquesta part de l'avaluació es basarà en el seguiment de la matèria. L'avaluació contínua representa el 30% de la nota final. Per la seua pròpia naturalesa, les proves i/o les activitats de l'avaluació contínua **NO SÓN RECUPERABLES**.

La nota final serà la suma ponderada de l'examen de síntesi i de l'avaluació contínua. En el cas que no s'aprove l'examen de síntesi (3,5 punts sobre 7, o alternativament 5 punts sobre 10), la nota final no podrà superar el màxim de 4,5 punts.

L'estudiant que no participe en l'avaluació contínua podrà ser avaluat en l'examen de síntesi i podrà obtenir una nota final màxima de 7 punts. Per aprovar l'assignatura l'estudiant ha d'obtenir un mínim de 5 punts sobre 7 (o alternativament 7 punts sobre 10) en l'examen de síntesi.

REFERÈNCIES



Bàsiques

- Baumer, B.S.; Kaplan, D.T. y Horton, N.J. (2017) Modern Data Science with R. CRC Press.
- CEACES, Proyecto (Contenedor Hipermedia de Estadística Aplicada a las Ciencia Económicas y Sociales). Universitat de València. ON LINE: <http://www.uv.es/ceaces>
- Esteban, J. y otros (2006). Estadística Descriptiva y nociones de probabilidad. Paraninfo, Madrid.
- Murgui, S. y otros (2002). Ejercicios de Estadística. Economía y Ciencias Sociales. Valencia: Tirant lo Blanch.
- Peng, R.D. (2015). Exploratory Data Analysis with R. Lean publishing. (bookdown: <https://bookdown.org/rdpeng/exdata/>)
- Teetor, P. (2011). R Cookbook. OREILLY
- Wichkam, H. y Golemund, G. (2016). R for Data Science. OREILLY.

Complementàries

- Newbold, P. y otros (2008). Estadística para Administración y Economía. Madrid Pearson-Prentice Hall, (6ª Edición).
- Pearson, R. K. (2018). Exploratory Data Analysis using R. Chapman and Hall/CRC.