

**Anàlisi Exploratòria de Dades**

Codi: 104853

Crèdits: 6

| Titulació                    | Tipus | Curs | Semestre |
|------------------------------|-------|------|----------|
| 2503852 Estadística Aplicada | FB    | 1    | 1        |

**Professor/a de contacte**

Nom: Rosa Camps Camprubi

Correu electrònic: rosa.camps@uab.cat

**Utilització d'idiomes a l'assignatura**

Llengua vehicular majoritària: català (cat)

Grup íntegre en anglès: No

Grup íntegre en català: Sí

Grup íntegre en espanyol: No

**Equip docent**

Montserrat Ferre Delgado

**Prerequisits**

Coneixements estàndard de Batxillerat.

**Objectius**

L'objectiu d'aquest curs és, en primer lloc, el coneixement de les tècniques descriptives i exploratòries aplicades a resumir la informació continguda en els conjunts de dades experimentals. Es comença amb l'estudi d'una variable, i seguidament s'inicia l'estudi comparatiu de dues o més variables. Es dediquen apartats especials a l'estudi de l'evolució temporals de les variables: sèries cronològiques i nombres índex. Un objectiu paral·lel, tant o més important, és la interpretació dels resultats i de les gràfiques en el context de les dades. Finalment, en aquesta assignatura s'aprèn el llenguatge i programari lliure R, de manipulació i anàlisi de dades, el qual s'utilitzarà al llarg del grau.

**Competències**

- Analitzar dades mitjançant l'aplicació de mètodes i tècniques estadístiques, treballant amb dades de diverses tipologies.
- Avaluar de manera crítica i amb criteris de qualitat el treball realitzat.
- Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
- Que els estudiants sàpiguin aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- Resumir i descobrir patrons de comportament en l'exploració de les dades.

- Seleccionar les fonts i tècniques d'adquisició i gestió de dades adequades per a fer-ne un tractament estadístic.
- Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
- Utilitzar correctament un ampli espectre del programari i llenguatges de programació estadístiques, escollint el més apropiat per a cada anàlisi i ser capaç d'adaptar-lo a noves necessitats.
- Utilitzar eficaçment la bibliografia i els recursos electrònics per obtenir informació.

## Resultats d'aprenentatge

1. Avaluar de manera crítica i amb criteris de qualitat la feina feta.
2. Depurar les dades: dades perdudes, transformació de variables, dades anòmales, selecció de casos i altres tècniques prèvies a l'anàlisi estadística.
3. Descriure, amb els mètodes gràfics i analítics adequats, dades de tipus qualitatiu en una o més variables.
4. Descriure, amb els mètodes gràfics i analítics adequats, dades de tipus quantitatiu en una o més variables.
5. Dissenyar modificacions de sintaxi en els programes per fer nous processos.
6. Explorar patrons de comportament de dades bivariades.
7. Explorar patrons de comportament de dades univariades.
8. Identificar i seleccionar les fonts d'informació més importants per a l'anàlisi descriptiva de dades de diferents tipologies: socials, ambientals, sanitàries, econòmiques, etc.
9. Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
10. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
11. Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
12. Treballar cooperativament en un context multidisciplinari assumint i respectant el rol dels diferents membres de l'equip.
13. Utilitzar eficaçment bibliografia i recursos electrònics per obtenir informació.
14. Utilitzar programes estadístics específics per a l'anàlisi descriptiva de dades.

## Continguts

### 1. Preliminars

- 1.1. Objectiu de l'anàlisi exploratòria de dades.
- 1.2. Tipus de variables i escales de mesura.
- 1.3. Arrodoniments i notació científica.

### 2. Resum de dades estadístiques.

- 2.1. Distribucions de freqüències: taules.
- 2.2. Agrupament en intervals.
- 2.3. Representacions gràfiques.

### 3. Característiques numèriques d'una variable.

- 3.1. Característiques de posició central: mitjana, mediana, moda.
- 3.2. Altres característiques de posició: quartils, decils i percentils.
- 3.3. Característiques de dispersió: variància i desviació típica (mostrals i poblacionals), rang, rang inter-quartil.
- 3.4. Característiques de dispersió relativa.
- 3.5. Puntuacions tipificades.
- 3.6. Característiques de forma: simetria i curtosi.

#### 4. Complements en l'estudi d'una variable.

- 4.1. Anàlisi exploratòria: diagrama de caixa i altres gràfics.
- 4.2. Transformació de variables.
- 4.3. Altres mitjanes: geomètrica, harmònica, quadràtica.
- 4.4. La desigualtat de Txebeixev.

#### 5. Comparació d'una variable en dos o més grups: Anàlisi exploratòria.

- 5.1. Situació de mostres independents.
- 5.2. Situació de mostres aparellades

#### 6. Tabulació i representació de la distribució conjunta dels valors de dues variables categòriques.

- 6.1. Taules de contingència (distribucions de freqüències conjunta, marginals i condicionades).
- 6.2. Anàlisi descriptiu de la dependència entre dues variables categòriques.

#### 7. Descripció numèrica de la distribució conjunta de dues variables estadístiques.

- 7.1. Característiques marginals i condicionades.
- 7.2. Corbes de regressió i coeficient de correlació.
- 7.3. Ajustament lineal i predicció.

#### 8. Introducció a les sèries temporals.

- 8.1. La descomposició clàssica.
- 8.2. Suavització de sèries: aplicació de filtres.

## Metodologia

El treball d'aula, teoria i problemes, es complementarà amb pràctiques d'ordinador on s'utilitzarà el paquet estadístic R.

Al campus virtual els alumnes hi trobaran la planificació de l'assignatura, les llistes de problemes i les pràctiques, així com eventuais canvis d'aula, horaris, etc.

És important que tingueu en compte que el CampusVirtual no és una web estàtica sinó que s'anirà actualitzant al llarg del curs.

Es procurarà introduir a la part més pràctica de l'assignatura, l'anàlisi i comparació de dades estadístiques per sexe, comentant, si s'escau, a l'aula les causes i els mecanismes socials i culturals que poden sustentar les desigualtats observades.

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

## Activitats formatives

| Títol                | Hores | ECTS | Resultats d'aprenentatge           |
|----------------------|-------|------|------------------------------------|
| Tipus: Dirigides     |       |      |                                    |
| Classes de problemes | 8     | 0,32 | 2, 3, 4, 6, 7, 8, 9, 10, 11        |
| Classes de teoria    | 18    | 0,72 | 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 13 |

|                                                               |    |      |                                               |
|---------------------------------------------------------------|----|------|-----------------------------------------------|
| Estudiar conceptes de teoria, resolver problemes a mà i amb R | 84 | 3,36 | 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14     |
| Pràctiques amb ordinador                                      | 30 | 1,2  | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 |

## Avaluació

La nota final de l'assignatura F s'obté a partir de:

- 1) Les notes dels dos exàmens parcials de teoria i problemes, TP1 i TP2, amb pesos respectius 20% i 25%.
- 2) Les notes de les dues proves amb ordinador, O1 i O2, amb pesos respectius 15% i 25%.
- 3) L'assistència a les sessions pràctiques amb ordinador i lliuraments que es proposin, PC, amb un pes del 15%. Aquesta part no és recuperable.

La nota final de l'assignatura s'obté fent la mitjana ponderada  $F = 0,2 TP1 + 0,15 O1 + 0,25 TP2 + 0,25 O2 + 0,15 PC$ .

*Requisit per superar l'assignatura amb la mitjana anterior*, les notes TP1, TP2 i O2 han de ser majors o iguals a 4 i que O1 sigui major o igual que 3,5.

Hi haurà dues proves de síntesi per als estudiants que no aprovin l'assignatura amb la fórmula anterior ( $F < 5$ ) o no compleixin els requisits mencionats (almenys 4 en O2, TP1, TP2 i almenys 3,5 en O1):

- Una prova de síntesi STP de tot el temari de teoria i problemes, per als estudiants que tinguin alguna de les notes TP1 o TP2 amb menys de 4 o suspenguin l'assignatura a causa de teoria i problemes.
- Una prova de síntesi SO de totes les pràctiques, per a aquells estudiants que tinguin menys de 3,5 en O1 i/o menys de 4 en O2 o bé suspenguin l'assignatura a causa de les pràctiques.

La seva nota final serà  $F = 0,45 STP + 0,40 SO + 0,15 PC$

(en cas que només els calgui presentar-se a una de les dues proves de síntesi se substituirà la nota de l'examen de síntesi per la mitjana ponderada dels dos parcials ja aprovats).

Si un estudiant no es presenta a cap parcial ni a la prova de síntesi, esqualificarà com a "No Avaluable".

Sense perjudici d'altres mesures disciplinàries que s'estimin oportunes, i d'acord amb la normativa acadèmica vigent, es qualificaran amb un zero les irregularitats comeses per l'estudiant que puguin conduir a una variació de la qualificació d'un acte d'avaluació. Per tant, plagiar, copiar o deixar copiar una pràctica o qualsevol altra activitat d'avaluació implicarà suspendre-la amb un zero i no es podrà recuperar en el mateix curs acadèmic. Si aquesta activitat té una nota mínima associada, aleshores l'assignatura quedarà suspesa.

Després de les segones proves parcials s'atorgaran les matrícules d'honor que es considerin clares. Aquestes matrícules seran ja definitives. Si el nombre màxim de matrícules permès no s'ha assolit, es reconsiderarà la possibilitat d'atorgar-ne més després de l'examen de recuperació.

## Activitats d'avaluació

| Títol                                            | Pes | Hores | ECTS | Resultats d'aprenentatge                   |
|--------------------------------------------------|-----|-------|------|--------------------------------------------|
| Lliuraments d'exercicis realitzats amb ordinador | 10% | 2     | 0,08 | 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 |
| Primer parcial de teoria i problemes             | 20% | 2     | 0,08 | 3, 4, 6, 7, 8, 9, 10, 11                   |

|                                                       |     |   |      |                                            |
|-------------------------------------------------------|-----|---|------|--------------------------------------------|
| Primera prova de resolució de problemes amb ordinador | 20% | 2 | 0,08 | 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14     |
| Segon parcial de teoria i problemes                   | 25% | 2 | 0,08 | 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 |
| Segona prova de resolució de problemes amb ordinador  | 25% | 2 | 0,08 | 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14     |

## Bibliografia

Manual de teoria

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

Bibliografia:

A.J.B. ANDERSON, Interpreting Data. A first course in Statistics, Ed Chapman and Hall, 1989.  
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.  
R. SAFRIAN, Introduction to Data Science, <https://bookdown.org/ronsarafian/IntrotoDS/>  
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.  
H. WICKHAM i G. GROLEMUND, R for Data Science, traducció a l'espanyol a <https://es.r4ds.hadley.nz/>  
R Tutorial. An R introduction to statistics. (2016). [www.r-tutor.com](http://www.r-tutor.com)

Bibliografia complementària.

G. CALOT, Curso de Estadística Descriptiva. Ed Paraninfo, 1988.  
FERNÁNDEZ, J.M. CORDERO, A. CORDERO, 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 volums)  
I. SOLER RAMOS i J. TORRENT SELLENS, Tècniques d'anàlisi de dades per a l'empresa, setembre 2011, Open Access UOC. <http://hdl.handle.net/10609/55081>

## Programari

R i RStudio, amb paquets bàsics i Tidyverse.