

# Guia docent

## 270950 - SIM - Inferència Estadística i Modelització

Última modificació: 23/11/2023

**Unitat responsable:** Facultat d'Informàtica de Barcelona  
**Unitat que imparteix:** 715 - EIO - Departament d'Estadística i Investigació Operativa.

**Titulació:** MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura obligatòria).

**Curs:** 2023      **Crèdits ECTS:** 6.0      **Idiomes:** Anglès

### PROFESSORAT

---

**Professorat responsable:** LIDIA MONTERO MERCADÉ

**Altres:** Primer quadrimestre:  
JOSEP FRANQUET FÀBREGAS - 12  
LIDIA MONTERO MERCADÉ - 11, 12

### CAPACITATS PRÈVIES

---

Els estudiants han de tenir coneixements suficients d'àlgebra i anàlisi matemàtica per assimilar els conceptes relacionats amb l'àlgebra de conjunts, sèries numèriques, funcions de variables reals d'una o més dimensions, derivació i integració. Els alumnes han d'haver cursat un curs de probabilitat i estadística

### COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

---

#### Específiques:

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

#### Genèriques:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

#### Transversals:

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

#### Bàsiques:

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB9. Que els estudiants posseïxin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

## METODOLOGIES DOCENTS

L'aprenentatge de l'assignatura consta de tres fases diferenciades:

1. Adquisició dels coneixements específics mitjançant l'estudi de la bibliografia i del material proporcionat pels professors.
2. L'adquisició de les destreses en les tècniques específiques d'anàlisi de dades, selecció del procés de modelatge estadístic i validació del model i
3. Integració dels coneixements, destreses i competències (específiques i transversals) mitjançant la resolució de casos d'estudi reals.

En les classes de Teoria s'exposen els fonaments de les metodologies i tècniques pròpies de l'assignatura. Les classes de laboratori serveixen per aprendre la utilització de les tècniques específiques per a la resolució de problemes, utilitzant les eines informàtiques adients, en aquest sentit els alumnes hauran primer de repetir un problema solucionat pels professors i després solucionar-ne un de similar al primer. Mentre que el Casos d'Estudi, resolt en grups i en hores bàsicament d'autoprenentatge, serveixen per posar en pràctica els coneixements, destreses i competències en la resolució de casos reals.

## OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Saber realitzar processos de inferència basats en les dades i de forma paramètrica tradicional per la presa de decisions.
2. Saber fer un informe sobre la qualitat de les dades i pre-processat
3. Determinació de les característiques significatives adreçades a targets numèrics i categòrics en grups d'individus
4. Estimació dels paràmetres i interpretació dels models lineals de resposta normal
5. Validació dels models de resposta normal. Identificació de dades inusuals i influents. Anàlisi dels residus
6. Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta normal
7. Estimació dels paràmetres i interpretació dels models lineals de resposta binària
8. Validació dels models de resposta binària. Identificació de dades inusuals i influents. Tipologia dels residus
9. Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta binària
10. Estimació dels paràmetres i interpretació dels models lineals de resposta polinòmica nominal i ordinal
11. Validació dels models de resposta polinòmica nominal i ordinal. Identificació de dades inusuals i influents.
12. Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta polinòmica nominal i ordinal
13. Estimació dels paràmetres i interpretació dels models lineals per comptatges
14. Validació dels models de per comptatges. Identificació de dades inusuals i influents. Tipologia dels residus. Diagnosi sobre dispersió. Models probabilístics paramètrics
15. Inferència d'hipòtesis sobre paràmetres simples i múltiples en models per comptatges
16. Saber fer un disseny d'experiments factorial i factorial fraccional

## HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup petit	27,0	18.00
Hores aprenentatge autònom	96,0	64.00
Hores grup gran	27,0	18.00

**Dedicació total:** 150 h

## CONTINGUTS

### Inferència clàssica versus fisheriana

#### Descripció:

Inferència clàssica. Funció de versemblança. Propietats de MLE. Prova de la ràtio de versemblança.

Procediments inferencials paramètrics vs no paramètrics.

Ús de dades històriques per a la prova d'hipòtesis. Relació amb la inferència fisheriana.



### Qualitat de les dades

**Descripció:**

Outliers univariants i multivariants.  
Dades mancants. Procediments d'imputació: deterministes, estocàstics.

### Models lineals normals

**Descripció:**

Descripció del model lineal normal. Estimació per mínims quadrats. Comparació de models. Deessa de l'estar en forma.  
Diagnòstic: dades influents i valors atípics. Ús de variables explicatives categòriques. Selecció de models. Predicció.  
Estimació de la xarxa neuronal de models de regressió lineal.

### Models lineals generalitzats

**Descripció:**

Enunciat dels models lineals generalitzats. Models per a dades de resposta binària. Models per a dades de recompte. Problemes de sobredispersió. Dades de resposta multinomial. Comparació de models. Diagnòstic: dades influents i valors atípics.  
Comparació i selecció de models.

### Disseny d'experiments

**Descripció:**

Disseny experimental factorial i factorials fraccional.  
Modernes tècniques d'anàlisi de dades per al disseny experimental

## ACTIVITATS

### Inferència clàssica versus fisheriana

**Descripció:**

Saber diferenciar les condicions d'aplicabilitat dels diferents mètodes d'inferència i saber escollir la més adequada al procés de Ciència de Dades entre mans.

Realitzar processos d'inferència per generar conclusions sobre les poblacions. Utilitzeu els valors  $p$ , els intervals de confiança i les proves per permutacions per a la presa de decisions i la interpretació de l'anàlisi en un problema de Ciència de Dades recurrents o puntuals.

**Objectius específics:**

1

**Competències relacionades:**

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 18h

Grup gran/Teoria: 4h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 12h

### Qualitat de les dades

**Descripció:**

Problemes en la qualitat de les dades: Es tracta de veure en el Cas d'Estudi els problemes que presenten o poden presentar les dades: Inconsistències, redundància. Dades mancants. Outliers. Com es fa un Informe de qualitat de les dades. En que consisteix l'estandardització de les dades.

**Objectius específics:**

2

**Competències relacionades:**

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 7h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 3h



### Determinació de las característiques significatives adreçades a targets numèrics i categòrics en grups d'individus

**Descripció:**

Aplicació de la inferència estadística per determinar les relacions entre las variables presents en una BBDD i una variable de resposta (numèrica o categòrica)

**Objectius específics:**

3

**Competències relacionades:**

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents  
CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

**Dedicació:** 5h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 1h

### Estimació dels paràmetres i interpretació dels models lineals de resposta normal

**Descripció:**

Perspectiva del modelatge per tècniques de regressió lineal : components estadístiques implicades. Rols: variables de resposta/explicatives. Estimació per mínims quadrats. Propietats dels estimadors. Processos inferencials involucrats.

**Objectius específics:**

4

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 8h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 4h



### Validació dels models de resposta normal. Identificació de dades inusuals i influents. Anàlisi dels residus

**Descripció:**

Elements que intervenen en la validació del modelatge per regressió. Valors influents i/o atípics

**Objectius específics:**

5

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 8h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 4h

### Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta normal

**Descripció:**

Inferència sobre els estimadors dels paràmetres en models lineals de resposta normal. Interval·ls de confiança, regions de confiança. Contrastos de hipòtesis simples, múltiples, combinacions lineals. Inferència sobre les prediccions i càlculs d'interval·ls de confiança.

**Objectius específics:**

6

**Competències relacionades:**

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 2h

### Estimació dels paràmetres i interpretació dels models lineals de resposta binària

**Descripció:**

Estimació màxim versemblant. Rol de la funció d'enllaç. Funció d'enllaç emprades. Propietats dels estimadors. Processos inferencials involucrats.

**Objectius específics:**

7

**Competències relacionades:**

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

**Dedicació:** 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 2h

### Validació dels models de resposta binària. Identificació de dades inusuals i influents. Tipologia dels residus

**Objectius específics:**

8

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 6h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 2h

### Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta binària

**Descripció:**

Inferència sobre estimadors de paràmetres en models lineals d'una resposta binària. Intervals de confiança. Contrastos d'hipòtesis múltiples i simples, combinacions lineals. Inferència sobre prediccions i càlculs d'intervals de confiança.

**Objectius específics:**

9

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

**Dedicació:** 3h

Grup gran/Teoria: 1h

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Estimació dels paràmetres i interpretació dels models lineals de resposta polinòmica nominal i ordinal

**Descripció:**

Estimació de màxima versemblança. Modelatge nominal versus ordinal. Funcions d'enllaç utilitzades. Propietats dels estimadors. Processos inferencials implicats.

**Objectius específics:**

10

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 4h

Grup gran/Teoria: 1h

Grup petit/Laboratori: 1h

Aprenentatge autònom: 2h





### Validació dels models de resposta politòmica nominal i ordinal. Identificació de dades inusuals i influents

**Descripció:**

Residus de la deviança, Pearson. Residus estudentitzats. Indicadors de dades inusuals i influents, mitjançant l'extensió dels indicadors emprats en la regressió normal.

**Objectius específics:**

11

**Competències relacionades:**

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents  
CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 2h 30m

Grup gran/Teoria: 0h 30m

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Inferència d'hipòtesis sobre paràmetres simples i múltiples en models de resposta politòmica nominal i ordinal

**Descripció:**

Inferència sobre estimadors de paràmetres en models lineals de resposta politòmica. Intervals de confiança. Contrastos d'hipòtesis simples, múltiples, combinacions lineals. Inferència sobre prediccions i càlculs d'intervals de confiança.

**Objectius específics:**

12

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents  
CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

**Dedicació:** 3h

Grup gran/Teoria: 1h

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Estimació dels paràmetres i interpretació dels models lineals per comptatges

**Descripció:**

Estimació de màxima versemblança. Modelatge Poisson, binomial negatiu. Sobredispersió. Funcions d'enllaç utilitzades. Processos inferencials involucrats.

**Objectius específics:**

13

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

**Dedicació:** 2h 30m

Grup gran/Teoria: 0h 30m

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Validació dels models de per comptatges. Identificació de dades inusuals i influents. Tipologia dels residus. Diagnosi sobredispersió. Models probabilístics paramètrics

**Descripció:**

Indicadors de dades poc habituals i influents. Comprovació de sobredispersió. Com superar la sobredispersió.

**Objectius específics:**

14

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 2h 30m

Grup gran/Teoria: 0h 30m

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Inferència d'hipòtesis sobre paràmetres simples i múltiples en models per comptatges

**Descripció:**

Inferència sobre estimadors de paràmetres en models lineals per comptatges. Interval de confiança. Contrastos d'hipòtesis simples, múltiples, combinacions lineals. Inferència sobre prediccions i càlculs d'interval de confiança.

**Objectius específics:**

15

**Competències relacionades:**

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

**Dedicació:** 2h 30m

Grup gran/Teoria: 0h 30m

Grup petit/Laboratori: 1h

Aprenentatge autònom: 1h

### Teoria i pràctica del disseny d'experiments factorial i factorial fraccional

**Objectius específics:**

16

**Competències relacionades:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdridigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 8h

Grup gran/Teoria: 2h

Grup petit/Laboratori: 2h

Aprenentatge autònom: 4h



## Examen Parcial

### Objectius específics:

1, 2, 3, 4, 5, 6

### Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

### Dedicació: 8h

Activitats dirigides: 2h

Aprenentatge autònom: 6h

## Examen Final

### Objectius específics:

7, 8, 9, 10, 11, 12, 13, 14, 15, 16

### Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

### Dedicació: 10h

Activitats dirigides: 2h

Aprenentatge autònom: 8h



### Pràctica de Models Lineals

#### Objectius específics:

2, 3, 4, 5, 6

#### Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 20h

Aprenentatge autònom: 20h

### Pràctica de Models Lineals Generalitzats

#### Objectius específics:

7, 8, 9, 10, 11, 12, 13, 14, 15

#### Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CG2. Identificar i aplicar mètodes d'anàlisi, extracció de coneixement i visualització de dades recollides en formats molt diferents

CE6. Dissenyar el procés de Ciència de Dades i aplicar metodologies científiques per a obtenir conclusions sobre poblacions i prendre decisions en conseqüència, a partir de dades estructurades o no estructurades i potencialment emmagatzemades en formats heterogenis.

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdirigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

**Dedicació:** 20h

Aprenentatge autònom: 20h

## SISTEMA DE QUALIFICACIÓ

---

L'avaluació de l'assignatura integra les tres fases d'aprenentatge descrites: coneixements, destreses i competencies.

Els coneixements s'avaluen mitjançant dos examens realitzats a la meitat (T1, pes 1/3) i durant la setmana d'exàmens del curs (T2, pes 2/3). En cas de suspendre l'examen parcial, l'alumne podrà repetir-lo com a extensió de l'examen final (nota T).

Les destreses s'avaluaràn a partir de l'entrega de 2 practiques, tantmateix com les competències transversals. Cadascun dels blocs 1, 2 i 3 per la primera pràctica (P1) i 4 i 5 per la segona (P2) comportaran una pràctica que l'alumne haurà de realitzar individualment o en grups de 2. El promig de les notes dona la nota P.

La Nota Final (NF) es calcula:

Examen Parcial (T1, 1/3) i Examen Final (T2, 2/3).

Pràctica 1 (P1) i Pràctica 2 (P2)

P: Nota de Pràctiques  $P=(P1+P2)/2$ .

T: Nota Teoria =  $\text{Max}(T2, (T1+2T2)/3)$ .

NF: Nota Final =  $0.6T + 0.4P$ .

## BIBLIOGRAFIA

---

### Bàsica:

- Fox, J.; Weisberg, S. An R companion to applied regression. 3rd ed. Thousand Oaks, California: SAGE Publications, Inc, 2019. ISBN 9781544336473.
- Wickham, H. Ggplot2: elegant graphics for data analysis. 2nd ed. Cham: Springer, 2016. ISBN 9783319242774.
- Montgomery, D. Design and Analysis of Experiments. 10th ed. Hoboken, N.J: Wiley, 2020. ISBN 9781119722106.
- Box, George E. P; Hunter, J. Stuart; Hunter, William Gordon. Statistics for experimenters : design, innovation, and discovery. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2005. ISBN 9780471718130.
- Fox, John. Applied regression analysis and generalized linear models. 3rd edition. Los Angeles: SAGE, 2016. ISBN 9781452205663.

### Complementària:

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, cop. 2009. ISBN 9780387952840.
- Trivedi, K.S,. Probability and statistics with reliability, queuing and computer science applications. 2nd ed. New York [etc.]: John Wiley and Sons, 2016. ISBN 1119285429.
- Mendenhall, W.; Wackerly, D.; Scheaffer, R. Mathematical Statistics with applications. Belmont, CA: Thomson Brooks/Cole, 2008. ISBN 9780495110811.

## RECURSOS

---

### Enllaç web:

- <https://www.r-project.org/>- <https://www.rstudio.com/>