
Pla docent

Nom de l'assignatura Anàlisi multivariant

Codi de l'assignatura M0165

Crèdits 5

Presentació

Descripció

L'Anàlisi Multivariant és un conjunt de mètodes estadístics i matemàtics per analitzar, descriure i interpretar les observacions multidimensionals, és a dir, el material estadístic que prové de l'observació de més d'una variable. A causa de les nombroses aplicacions que té en la pràctica totalitat de les ciències experimentals, l'Anàlisi Multivariant ha tingut un desenvolupament creixent en els últims anys i la seva utilització s'ha convertit si fa no fa en imprescindible. Les espectaculars possibilitats que actualment ofereix la informàtica, tant en programari com en maquinari, han influït de forma decisiva en aquest desenvolupament.

El nombre actual de mètodes d'Anàlisi Multivariant és massa elevat per poder-los tractar dins dels límits d'una assignatura. Per tant s'han seleccionat aquells més consolidats i de major aplicació.

Camps professionals en què es projecta

Totes les ciències experimentals.

Coneixements previs

Encara que farem un repàs d'alguns temes, és convenient que l'alumnat revisi el càlcul matricial, la geometria lineal i les principals distribucions i contrastos univariants.

Objectius i competències

Pel que fa a coneixements:

L'objectiu principal d'aquesta assignatura és que l'alumnat es familiaritzi amb algunes de les tècniques més comuns de l'anàlisi multivariant.

Bàsicament, tindrem en compte tres grans blocs de tècniques:

- 1) tècniques de representació i de reducció de la dimensió,
- 2) inferència multivariant,
- 3) tècniques de classificació: anàlisi de conglomerats i anàlisi discriminant.

L'assignatura té un plantejament aplicat. Essencialment s'espera que l'alumnat aprengui:

- Quines són les tècniques oportunes per a cada situació (saber "què es pot fer").
- Quines són les restriccions que s'apliquen en cada cas, és a dir, quan es pot aplicar i quan no es pot aplicar una tècnica determinada (saber "si es pot fer").
- Com s'ha d'aplicar cada tècnica (saber "com s'ha de fer").

Donada l'orientació pràctica de l'assignatura, es concedirà molta importància a la utilització de l'ordinador mitjançant el programari adequat per aplicar tècniques ja implementades o per programar alguns algorismes senzills. En concret utilitzarem el llenguatge estadístic R i els seus paquets en aquest camp.

Un altre objectiu és generalitzar alguns conceptes, que segurament l'alumnat ja coneix en el cas univariant o uniparamètric, en una formulació més general. Això fa que sigui molt recomanable tenir un bon bagatge de coneixements de Inferència Estadística, com la proporcionada per l'assignatura del mateix nom, així com d'Àlgebra Matricial, que es revisarà a l'inici de l'assignatura.

Continguts

Repte 1. Introducció a l'anàlisi multivariant.

- Descripció de dades multivariants
 - La matriu de dades: objectes i variables. Anàlisi univariant.
 - Mesures de centralització.
 - La matriu de variàncies i covariàncies.

- Mesures globals de variabilitat.
- Distàncies.
- Anàlisi gràfica i dades atípiques.
 - Representacions gràfiques. Histogrames. Diagrames de dispersió.
 - Diagrames de caixes múltiples i altres gràfics multidimensionals.
 - Dades atípiques. Els seus efectes. Identificació.
- Elements d'àlgebra matricial i càlcul matricial.
 - Càlcul matricial.
 - Diagonalització.
 - Descomposició en valors singulars.
- *Aplicació: Estudi exploratori d'una base de dades.*
- *Aplicació: Anàlisi de correspondències múltiple: càlcul matricial.*

Repte 2. Anàlisi de components principals.

- Plantejament del problema. Objectius.
- Càlcul de les components.
- Propietats de les components.
- Representació gràfica dels objectes.
- Interpretació de les components.
- Elecció de l'nombre de components.
- Utilització de les components principals en regressió.
- *Aplicació: Detecció d'efectes "batch" en dades de microarrays*

Repte 3. Anàlisi de proximitats (multidimensional scaling)

- Definició, propietats i relacions entre similituds i dissimilituds.
- Distàncies per a variables binàries, categòriques, ordinals i mixtes.
- Distàncies euclidianes.
- Anàlisi de coordenades principals.

- Introducció al MDS no mètric.

Repte 4. Anàlisi de correspondències.

- Plantejament del problema.
- El llenguatge: perfils, masses, centroide, distància chi-quadrat.
- Projectió de les files.
- Projectió de les columnes.
- Tipus de representació simultània.

Repte 5. Estadística matemàtica multivariant.

- Vectors aleatoris i distribucions multivariants.
- Funció de distribució i funció de densitat. Funcions marginals i condicionades.
- Vector d'esperances, matriu de covariàncies i matriu de correlacions.
- La distribució multinomial. (opcional)
- La distribució normal multivariant.
- Estimació i inferència estadística multivariant.
 - Estimacions de màxima versemblança.
 - Distribucions per a la inferència en poblacions normals multivariants.
- Contrast d'hipòtesis sobre la mitjana en una i dues poblacions normals.
- Contrast d'hipòtesis en k poblacions normals: MANOVA d'un factor.
- Contrastos sobre l'homogeneïtat de la variabilitat.
- Proves de normalitat multivariant
- Transformacions normalitzants. (opcional)
- Aplicació: *Comparació multivariant de dues poblacions.*

Repte 6. Anàlisi de conglomerats (cluster analysis)

- Objectius.
- Anàlisi jeràrquic de conglomerats. Dendogrames.
- Mètodes no jeràrquics: k-mitjanes i PAM.

- Nombre de classes. La silueta.
- *Aplicació: Anàlisi i visualització de patrons d'expressió gènica*

Repte 7. Anàlisi discriminant

- Entre dues poblacions. Cas normal: discriminador lineal i quadràtic.
- Cas multinomial. El criteri de la dispersió: el discriminador de Fisher.
- Criteri de la mínima distància.
- Taules de classificació. Validació creuada. Probabilitats *a posteriori*.
- Un mètode de classificació no paramètric: k-nearest neighbors.
- Regressió logística.
- *Aplicació: Classificació molecular amb dades de microarrays.*

Recursos

Informació addicional sobre els recursos d'aprenentatge i eines de suport

L'equip docent aportarà un vídeo de presentació de cada mòdul.

A més de la bibliografia, entre els recursos de cada mòdul hi ha tutorials específics en format PDF, scripts amb instruccions de R i vídeos.

Cada mòdul tindrà un bon conjunt d'exercicis, alguns opcionals i amb diferents nivells de dificultat.

Els marcats amb (*) seran opcionals.

Els marcats amb (**), a més d'opcionals, són de major dificultat.

Llista dels principals recursos d'aprenentatge

Material	Suport
Anàlisi de components principals	Audiovisual

Anàlisi de conglomerats	Audiovisual
Anàlisi de correspondències	Audiovisual
Anàlisi de proximitats	Audiovisual
Anàlisi discriminant	Audiovisual
Anàlisi multivariant: introducció	Audiovisual
Análisis de componentes principales	Audiovisual
Análisis de conglomerados	Audiovisual
Análisis de correspondencias	Audiovisual
Análisis de proximidades	Audiovisual
Análisis discriminante	Audiovisual
Análisis multivariante: introducción	Audiovisual
Inferencia en poblaciones normales multivariantes. Comparación de varianzas y test de normalidad	Audiovisual

Inferencia en poblaciones normales multivariantes. Contraste de dos medias	Audiovisual
Inferencia en poblaciones normales multivariantes. MANOVA	Audiovisual
Inferencia multivariante	Audiovisual
Inferència en poblacions normals multivariants. Comparació de variàncies i testos de normalitat	Audiovisual
Inferència en poblacions normals multivariants. Contrast de dues mitjanes	Audiovisual
Inferència en poblacions normals multivariants. MANOVA	Audiovisual
Inferència multivariant	Audiovisual
Teoria de l'anàlisi de correspondències	Audiovisual
Teoria de l'anàlisi de correspondències: exemple	Audiovisual
Teoría del análisis de correspondencia: ejemplo	Audiovisual
Teoría del análisis de correspondencias	Audiovisual

Informació addicional sobre la bibliografia i fonts d'informació

Bibliografia bàsica

MANLY, B. F. J. *Multivariate Statistical Methods: A Primer*. Fourth Edition. Londres: Chapman and Hall, 2016.

EVERITT, B.; HOTHORN, T. *An introduction to applied multivariate analysis with R*. Springer, 2011.

PEÑA, D. *Análisis de datos multivariantes*. Madrid. McGraw Hill. 2002

Bibliografia complementària

CUADRAS, C.M. *Métodos de análisis multivariante*. Barcelona: EUB, 1996.

CUADRAS, C.M. *Nuevos Métodos de Análisis Multivariante*. Barcelona: CMC Editions, 2018 (PDF)

EVERITT, B. S. *An R and S-PLUS Companion to Multivariate Analysis*. Londres: Springer-Verlag, 2005.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. *Multivariate Analysis*. Londres: Academic Press, 1979.

GREENACRE, M. J. La práctica del análisis de correspondencias. 2008 (PDFs en <https://www.fbbva.es/microsite/multivariate-statistics/practica.html>)

GREENACRE, M. J. *Correspondence analysis in practice*. 3rd Edition. Londres: Chapman-Hall, 2016.

JOHNSON, R. A.; WICHEM, D. W. *Applied Multivariate Statistical Methods*. London: Prentice Hall, 2002.

Metodologia

Abans de començar cada tema es subministrarà una guia d'estudi, on s'han d'assenyalar els materials de lectura corresponents a el tema. Els materials de lectura poden consistir en capítols o apartats de l'material docent aportat, bé en llibres o bé en tutorials, manuals o articles que es proporcionen en format PDF o per hiperenllaços.

La lectura d'el material didàctic seguint les orientacions de la guia conforma *la primera activitat de cada bloc* i que haurà de realitzar

l'alumnat de forma individual. Amb això es pretén que l'estudiant es familiaritzi amb el contingut de la matèria així com amb el material en si mateix, el qual es transformarà en material de consulta per a l'estudiant durant el desenvolupament de la lliçó. Lògicament, aquesta activitat s'ha de desenvolupar en els primers dies de l'bloc, per poder abordar posteriorment els exercicis, els possibles debats i les PAC.

La segona activitat serà resoldre els exercicis:

- Exercicis teòrics (pocs).
- Exercicis aplicats (els més): Exercicis amb dades i resultats numèrics i gràfics i la seva interpretació.

A més, és possible que s'introdueixin alguns debats sobre els conceptes i les seves propietats o sobre aspectes de la implementació pràctica d'alguns mètodes. Aquests debats han de ser el fòrum on es manifeste la comprensió dels aspectes centrals del tema o les dificultats que aquest comporta. També es poden introduir conceptes complementaris.

Avaluació

Informacions sobre l'avaluació a la UOC

A la UOC, l'avaluació generalment és virtual. S'estructura entorn de l'avaluació contínua, que inclou diferents activitats o reptes; l'avaluació final, que es porta a terme mitjançant proves o exàmens, i el treball final de la titulació.

Les activitats o proves d'avaluació poden ser escrites i/o audiovisuals, amb preguntes aleatòries, proves orals síncrones o asíncrones, etc., d'acord amb el que decideixi cada equip docent. Els treballs finals representen el tancament d'un procés formatiu que implica la realització d'un treball original i tutoritzat que té com a objectiu demostrar l'adquisició competencial feta al llarg del programa.

Per verificar la identitat de l'estudiant i l'autoria de les proves d'avaluació, la UOC es reserva la potestat d'aplicar diferents sistemes de reconeixement de la identitat i de detecció del plagi. Amb aquest objectiu, la UOC pot dur a terme enregistrament audiovisual o fer servir mètodes o tècniques de supervisió durant l'execució de qualsevol activitat acadèmica.

Així mateix, la UOC pot exigir a l'estudiant l'ús de dispositius electrònics (micròfons, càmeres o altres eines) o programari específic durant

L'avaluació. És responsabilitat de l'estudiant assegurar que aquests dispositius funcionen correctament.

El procés d'avaluació es fonamenta en el treball personal de l'estudiant i pressuposa l'autenticitat de l'autoria i l'originalitat de les activitats acadèmiques. Al [web sobre integritat acadèmica i plagi de la UOC](#) hi ha més informació respecte d'aquesta qüestió.

La manca d'autenticitat en l'autoria o d'originalitat de les proves d'avaluació; la còpia o el plagi; la suplantació d'identitat; l'acceptació o l'obtenció de qualsevol activitat acadèmica a canvi d'una contraprestació o no; la col·laboració, l'encobriment o l'afavoriment de la còpia, o l'ús de material, programari o dispositius no autoritzats en el pla docent o l'enunciat de l'activitat acadèmica, inclosa la intel·ligència artificial i la traducció automàtica, entre altres, són conductes irregulars en l'avaluació que poden tenir conseqüències acadèmiques i disciplinàries greus.

Aquestes conductes irregulars poden comportar el suspens (D/0) en les activitats avaluable que es defineixin en el pla docent -incloses les proves finals- o en la qualificació final de l'assignatura, sigui perquè s'han utilitzat materials, programari o dispositius no autoritzats durant les proves (com l'ús d'intel·ligència artificial no permesa, xarxes socials o cercadors d'informació a internet), perquè s'han copiat fragments de text d'una font externa (internet, apunts, llibres, articles, treballs o proves d'altres estudiants, etc.) sense la citació corresponent, per la compravenda d'activitats acadèmiques, o perquè s'ha dut a terme qualsevol altra conducta irregular.

Així mateix, i d'acord amb la [normativa acadèmica](#), les conductes irregulars en l'avaluació també poden donar lloc a la incoació d'un procediment disciplinari i a l'aplicació, si escau, de la sanció que correspongui, de conformitat amb el que estableix la [normativa de convivència](#) de la UOC.

En el marc del procés d'avaluació, la UOC es reserva la potestat de:

- Sol·licitar a l'estudiant que acrediti la seva identitat segons el que estableix la normativa acadèmica.
- Sol·licitar a l'estudiant que acrediti l'autoria del seu treball al llarg de tot el procés d'avaluació, tant en l'avaluació contínua com en l'avaluació final, per mitjà d'una entrevista oral síncrona, que pot ser objecte d'enregistrament audiovisual, o pels mitjans que estableixi la Universitat. Aquests mitjans tenen l'objectiu de verificar els coneixements i les competències que garanteixin la identitat de l'estudiant. Si no és possible garantir que l'estudiant és l'autor de la prova, aquesta pot ser qualificada

amb una D, en el cas de l'avaluació contínua, o amb un suspens, en el cas de l'avaluació final.

Intelligència artificial en el marc de l'avaluació

La UOC reconeix el valor i el potencial de la intel·ligència artificial (IA) en l'àmbit educatiu, alhora que posa de manifest els riscos que comporta si no s'utilitza de manera ètica, crítica i responsable. En aquest sentit, en cada activitat d'avaluació s'informarà l'estudiantat sobre les eines i els recursos d'IA que es poden utilitzar i en quines condicions. Per la seva banda, l'estudiantat es compromet a seguir les indicacions de la UOC a l'hora de dur a terme les activitats d'avaluació i de citar les eines utilitzades i, concretament, a identificar els textos o les imatges generats per sistemes d'IA, els quals no podrà presentar com si fossin propis.

Amb relació a fer servir o no la IA per resoldre una activitat, l'enunciat de les activitats d'avaluació indica les limitacions en l'ús d'aquestes eines. Cal tenir en compte que fer-les servir de manera inadequada, com ara en activitats en què no estan permeses o no citar-les en les activitats en què sí que ho estan, es pot considerar una conducta irregular en l'avaluació. En cas de dubte, es recomana que, abans de lliurar l'activitat, es faci arribar una consulta al professorat col·laborador de l'aula.

Model d'avaluació

L'assignatura només es pot aprovar amb el seguiment i la superació de l'avaluació contínua (AC). La qualificació final de l'assignatura és la nota obtinguda a l'AC.

Opció per superar l'assignatura: AC

Nota final d'assignatura: AC

Avaluació continuada

Per superar aquesta assignatura l'estudiant deu:

- Realitzar els exercicis obligatoris i opcionals proposats en cada mòdul. Aquesta activitat no tindrà una avaluació particular, encara que pot ajudar en la qualificació final.

- Participar al foro i als debats associats a cada tema, si es proposen, tot seguint les orientacions del consultor i que serviran per complementar la nota de cada PAC.
- Realitzar i lliurar en el dia fixat cadascuna de les PAC proposades al final dels blocs temàtics.
- En la qualificació final cadascuna de les PAC tindrà el mateix pes.

Es valorarà especialment:

- La compressió i relació dels conceptes treballats.
- La resolució dels exercicis proposats en cada mòdul.
- Capacitat de presentació i redacció dels conceptes i anàlisis contingudes en les PAC.
- Domini de l'argumentació on-line com a mecanisme de confrontació i creació de coneixement, demostrada a través de la participació en el fòrum i en els debats.

Les solucions dels exercicis es poden discutir en el fòrum. Les solucions definitives de cada estudiant es poden compartir després de la publicació de la solució del conjunt dels exercicis de cada tema per part del professorat.

Les solucions de les PAC específiques es faran públiques al mateix temps que les notes obtingudes. D'aquesta manera l'alumnat podrà contrastar les seves respostes amb les solucions correctes i així modificar idees errònies o incidir en l'estudi dels punts més fluixos.

La planificació proposada es desenvolupa al llarg de catorze setmanes (5 crèdits) en les quals l'alumnat ha d'ajustar-se per a un adequat seguiment de l'assignatura. Dins d'aquestes catorze setmanes es fixaran les dates de lliurament dels exercicis que poden variar lleugerament en funció del ritme del curs.

Per aprovar l'assignatura, els alumnes amb una qualificació de D o N en la PAC1 hauran de realitzar i superar una prova addicional. En aquesta prova addicional la nota màxima serà un 5. Els alumnes amb una qualificació C- en la PAC1 també podran realitzar aquesta prova per millor la nota final. La prova addicional s'entregarà a la mateixa data que la PAC2. La PAC2 no tindrà recuperació.

Una qualificació de D o N a la PAC2 implica no aprobar l'assignatura.

Avaluació final

Aquesta assignatura no té avaluació final.

Dates de les activitats

Nom de l'activitat	Data d'inici	Data de lliurament
Lliurament de l'activitat PAC 1	23/04/2024	07/05/2024
Lliurament de l'activitat PAC 2	15/06/2024	30/06/2024