



Universitat de Lleida

GUIA DOCENT

# **APLICACIONS DE L'ANÀLISI MULTIVARIANT EN BIOTECNOLOGIA**

Coordinació: GATIUS CORTIELLA, FERNANDO

Any acadèmic 2023-24

## Informació general de l'assignatura

Denominació	APLICACIONS DE L'ANÀLISI MULTIVARIANT EN BIOTECNOLOGIA			
Codi	101636			
Semestre d'impartició	1R Q(SEMESTRE) AVALUACIÓ CONTINUADA / 2N Q(SEMESTRE) AVALUACIÓ CONTINUADA			
Caràcter	Grau/Màster	Curs	Caràcter	Modalitat
	Grau en Biotecnologia	4	OPTATIVA	Presencial
Nombre de crèdits assignatura (ECTS)	6			
Tipus d'activitat, crèdits i grups	Tipus d'activitat	PRAULA		TEORIA
	Nombre de crèdits	3		3
	Nombre de grups	2		2
Coordinació	GATIUS CORTIELLA, FERNANDO			
Departament/s	QUÍMICA, FÍSICA, CIÈNCIES AMBIENTALS I DEL SÒL			
Distribució càrrega docent entre la classe presencial i el treball autònom de l'estudiant	60 hores presencials + 90 hores no presencials per grup (a cada grup s'imparteix exactament la mateixa docència)			
Informació important sobre tractament de dades	Consulteu <a href="#">aquest enllaç</a> per a més informació.			
Idioma/es d'impartició	Català			
Distribució de crèdits	L'assignatura s'imparteix ÍNTEGRAMENT en 2 blocs diferents d'optatives:  - GRUP 1 (màxim 12 estudiants): 6 CRÈDITS IMPARTITS AL BLOC 1 D'OPTATIVES (INICI AL SETEMBRE) - GRUP 2 (màxim 12 estudiants): 6 CRÈDITS IMPARTITS AL BLOC 3 D'OPTATIVES (INICI AL FEBRER)  IMPORTANT!!!: en el moment de la matrícula l'estudiant ha de seleccionar un dels dos grups degut a que cada grup està limitat a un NOMBRE MÀXIM DE 12 ESTUDIANTS.			

Professor/a (s/es)	Adreça electrònica professor/a (s/es)	Crèdits impartits pel professorat	Horari de tutoria/lloc
GATIUS CORTIELLA, FERNANDO	fernando.gatius@udl.cat	12	

## Informació complementària de l'assignatura

IMPORTANT: si alguna activitat no es pot desenvolupar amb normalitat es substituirà per una activitat alternativa o es realitzarà virtualment.

Dades de contacte del professor:

**Ferran Gatus Cortiella** (Departament de Química. Universitat de Lleida)

Despatx A 0.07.2. Pdís. Planta Baixa - Edifici A. Campus ETSEA. Av. Rovira Roure, 191. 25198 - Lleida

Telèfon núm. [+34 973 00 36 54](tel:+34973003654) ; E-mail: [gatius@quimica.udl.cat](mailto:gatius@quimica.udl.cat)

## Objectius acadèmics de l'assignatura

L'estudiant que superi l'assignatura ha de: (Objectius de coneixement)

- ENTENDRE i UTILITZAR la terminologia bàsica i els conceptes fonamentals que s'aborden en la matèria especificats en el programa de la mateixa.
- Saber utilitzar els conceptes mostrats per INTERPRETAR resultats obtinguts de l'anàlisi de dades de diferent naturalesa, tant en si estan caracteritzades per una o moltes variables.
- Aconseguir una formació de caire científic basada en el coneixement de la metodologia implicada en el tractament de dades, així com en la universalitat que confereix el caràcter APLICAT de l'assignatura a dades de diferents tipus.
- Més generalment, demostrar una preparació per al disseny d'investigacions sobre la base dels principis de rellevància, control i generalització.

L'estudiant que superi l'assignatura ha de ser capaç de: (Objectius de capacitat)

- PLANTEJAR i DESENVOLUPAR l'estudi i el tractament de diferents tipus de dades en funció del seu caràcter.
- Expressar correctament el resultat d'un estudi i realitzar comparacions de resultats numèrics utilitzant les eines i contrastos adequats a cada cas.
- Plantejar i ANALITZAR/INTERPRETAR els models implicats en l'estudi descriptiva d'un conjunt de dades dins l'àmbit multivariant per extreure'n les conclusions pertinents.
- ENTENDRE i APLICAR les tècniques multivariants de projecció per a altres estudis apart dels descriptius, com són els implicats en el càlcul de relacions entre diferents variables i els procediments de classificació de mostres.
- INTERPRETAR, DISCUTIR i EXTREURE CONCLUSIONS d'estudis multivariants de dades (en general), tan si han estat realitzats per un mateix com per una altra persona.

## Competències

El graduat en Biotecnologia ha de ser capaç de:

CG2 Interpretar la informació científicotècnica amb un sentit crític, i ser capaç de fer presentacions basades en aquesta informació.

CG6 Conèixer i saber utilitzar el programari i les bases de dades específiques en els diferents àmbits de la Biotecnologia.

CG7 Utilitzar el mètode científic per analitzar dades i dissenyar estratègies experimentals amb aplicacions biotecnològiques.

CG11 Adquirir criteris d'elecció de les tècniques analítiques més adequades per a cada cas pràctic concret.

## Continguts fonamentals de l'assignatura

En la major part d'estudis que es desenvolupen es treballa amb mostres/dades definides per moltes variables. El tractament d'aquesta informació és essencial per a extreure les conclusions dels estudis. En aquesta assignatura es presenten una sèrie d'eines per a poder tractar tota aquesta informació de forma eficient, SENSE la necessitat de coneixements avançats de matemàtiques o estadística, i es dona una visió global de l'anàlisi i tractament de dades amb aquestes eines i, sobre tot, de la seva APLICACIÓ per a l'estudi i extracció de conclusions.

Per a fer això es comença (a mode d'introducció) situant a l'alumne amb el repàs (molt breu) dels principals conceptes del tractament univariant de dades (que l'alumne ja coneix de primer curs) que li permetran arribar a l'expressió correcta del resultat d'un estudi i a fer tractaments bàsics dels valors numèrics que defineixen una variable. En aquesta introducció es repassen conceptes estadístics bàsics que es connecten amb la presentació i aplicació de les tècniques de projecció com a eines per a tractar dades de caire multivariant. Aquestes eines s'APLIQUEN als estudis descriptius de dades, a la relació entre variables i a la classificació d'individus (mostres, unitats experimentals) atenent a grups preestablerts.

Cal tenir en compte que és una assignatura d'anàlisi de dades APLICADA, en la que s'emprarà el software corresponent per a fer qualsevol càlcul/model dels implicats en un estudi, i en la que s'avaluarà la COMPRENSIÓ dels CONCEPTES implicats en la mateixa mitjançant la INTERPRETACIÓ dels fenòmens que mostren els diagrames generats en els esmentats càlculs/models (interpretació que conduirà a l'extracció de les conclusions pertinents). Són fonamentals les classes PRÀCTIQUES en les que s'ensenyarà tant a utilitzar l'esmentat software com a interpretar els models/diagrames implicats en els diferents estudis a mode de tutorial.

## PLANIFICACIÓ DELS CONCEPTES TEÒRICS QUE ES TRACTARAN EN L'ASSIGNATURA

### INTRODUCCIÓ: REPÀS DEL TRACTAMENT DE DADES UNIVARIANTS

Errors en l'anàlisi quantitativa de resultats. Tipus. Repetibilitat / reproduïbilitat

Estadístics per a les determinacions amb mesures repetides. Distribucions. Normalitat

Concepte de MOSTRA. La distribució de mitjanes

Límits de confiança. Aplicacions a la presentació dels resultats d'un estudi

Contrastos de significació: comparació de la precisió i el resultat de 2 mètodes

Comparació de més de 2 valors (estudis amb factors - ANOVA)

Altres contrastos: comparació de freqüències, detecció d'outliers

Determinació de la normalitat d'una distribució de mesures

Conclusions a partir dels contrastos de significació. Errors associats

### DEFINICIÓ DE L'ÀMBIT MULTIVARIANT PER A L'ESTUDI DE CASOS

Observacions indirectes i correlacions

Anàlisi Multivariant i Estadística Multivariant

Presentació de les dades multivariants: el conjunt (matriu) de dades objecte d'un estudi

Objectius principals de les tècniques d'anàlisis multivariant

## MÈTODES DE PROJECCIÓ. ANÀLISI PER COMPONENTS PRINCIPALS (PCA)

Estudi de dades definides per variables correlacionades. Plantejament i problemes

El primer Component Principal (PC): definició i obtenció. Extensió del model a més PCs

Conceptes de *scores* i *loadings* (projeccions/representació de mostres i variables)

Variància explicada i variància residual de les dades en un model PCA

Residual i *leverage* de les mostres de l'estudi

Estudi dels *outliers* o mostres estranyes que es poden trobar en un estudi

Interpretació del model PCA i dels diagrames associats

## RELACIÓ ENTRE VARIABLES. LA REGRESSIÓ MULTIVARIANT

Estudi de la regressió en àmbit univariant. Repàs dels principals conceptes de l'ajust per mínims quadrats per emprar-lo com a base de comparació en la regressió multivariant

Extensió de l'estudi a l'àmbit multivariant. La regressió lineal múltiple (MLR) Definició del model MLR. Limitacions i problemes dels models MLR. Primera solució als problemes trobats: la selecció de variables

Solució definitiva als problemes trobats: la Regressió per Components Principals (PCR). Necessitat dels Components Principals en l'àmbit de la regressió multivariant. Estudi del model: *scores*, *loadings*, variàncies, error. Limitacions/problemes dels models PCR

Solució als problemes de PCR: la Regressió per Mínim Quadrats Parcial (PLS) Definició dels factors PLS. Tipus de model PLS. Estudi dels models: representacions (*scores*, *loadings*) i errors associats

## LA VALIDACIÓ DELS MODELS MULTIVARIANTS

Calibratge i validació de models: requeriments dels conjunts associats

El concepte de "conjunt de test" en la validació de models

Altres mètodes: la correcció per importància (*leverage*) i la validació creuada

Estudi dels errors. Mesura de l'error experimental (estudis amb repeticions). Bondat dels models de regressió: RMSEP, RMSEC i altres paràmetres

## MÈTODES DE CLASSIFICACIÓ DINS L'ÀMBIT MULTIVARIANT

Classificacions supervisades i no supervisades

Aplicació dels mètodes de projecció a la classificació: el mètode SIMCA. Definició dels models que defineixen les classes

Definició de la metodologia de classificació. Paràmetres

Anàlisi dels resultats (taules i diagrames) del procediment de classificació

## Activitats pràctiques

1. Exercicis de repàs d'estudis univariants amb full de càlcul: contrastos d'hipòtesis, anàlisi de variància i regressió lineal.
2. Exploració de dades. Mètodes de projecció. Anàlisi per Components Principals (PCA). Representació de la informació (*scores* i *loadings*). Residuals i *leverage*.
3. Regressió per Components Principals (PCR). Paràmetres de calibratge. Distribucions de mostres i variables en el model obtingut. Error del model. Interpretació. Definició del model: obtenció dels coeficients de regressió.
4. Regressió per Mínims Quadrats Parcial (PLS). Detecció de mostres estranyes (*outliers*) i tractament de les mateixes en l'obtenció del model. Interpretació del model trobat: distribucions de mostres i variables. Error del model i obtenció dels coeficients de regressió. Comparació PCR-PLS. Comparació de diferents mètodes de Validació de models.
5. Mètodes de classificació. Classificació supervisada i no supervisada. Mètode SIMCA. Anàlisi dels resultats de la classificació relatius a objectes, variables i models de les classes.

## Eixos metodològics de l'assignatura

Tipus d'activitat	Descripció	Activitat presencial alumne		Activitat no presencial alumne		Temps total
		Objectius	Hores	Treball alumne	Hores	
<b>Lliçó magistral</b>	Classe magistral (Aula. Grup gran)	Explicació dels principals conceptes	30	Estudi: Conèixer, comprendre i sintetitzar coneixements	45	<b>75hores 3ECTS</b>
<b>Problemes i casos</b>	Classe participativa (Aula. Grup gran )	Resolució de problemes i casos (aplicació dels conceptes explicats, pel que es planteja juntament amb l'anterior)	-	Aprendre a resoldre problemes i casos (aplicar els coneixements a dades reals, pel que es planteja juntament amb l'anterior)	-	-
<b>Seminari</b>	Classe participativa (Grup mitjà)	Realització d'activitats de discussió o aplicació	20	Resoldre problemes i casos. Discutir	30	<b>50hores 2ECTS</b>
<b>Laboratori</b>	Pràctica de Laboratori (Grup mitjà)					
<b>Aula d'informàtica</b>	Pràctica a l'aula (Grup mitjà)	Execució de la pràctica: comprendre fenòmens, mesurar...	10	Estudiar, entendre i aplicar	15	<b>25hores 1ECTS</b>
<b>Pràctiques de camp</b>						
<b>Visites</b>						
<b>Activitats dirigides</b>						
<b>Altres</b>						
<b>Totals</b>			<b>60</b>		<b>90</b>	<b>150hores 6ECTS</b>

## Sistema d'avaluació

Durant el curs es van resolent i comentant a classe exercicis pràctics d'aplicació de les tècniques que s'expliquen a les sessions de caire més teòric.

**L'assignatura es divideix en tres blocs de continguts:** després de cadascun d'ells es realitzarà una prova d'avaluació consistent en exercicis pràctics d'aplicació com els que s'han resolt anteriorment a classe.

Cadascuna de les proves contarà un terç de la nota final de l'assignatura.

Si al final del curs l'estudiant no està d'acord amb la qualificació obtinguda en aquestes proves, tindrà dret a una convocatòria extraordinària per canviar aquesta nota.

Encara que és altament recomanable l'assistència a les classes, aquells estudiants **que puguin demostrar la impossibilitat d'assistència a classe durant el curs per motius laborals**, podran demanar a començament de curs (a la direcció d'estudis del centre) una avaluació alternativa de tota l'assignatura que tindrà lloc el mateix dia de l'avaluació del darrer bloc de continguts.

## Bibliografia i recursos d'informació

### Bibliografia bàsica

Multivariate data analysis in practice ESBENSEN, K.; SCHÖNKOPF, S.; MIDTGAARD, T. CAMO AS. Trondheim. Norway. 1996.

Chemometrics. Data Analysis for the Laboratory and Chemical Plant. BRERETON, R.G. John Wiley & Sons, Chichester. 2003.

Multivariate Calibration MARTENS, H.; NAES, T. John Wiley & Sons. Chichester. 1989.

### Bibliografia complementària

Statistics for analytical chemistry (3rd Ed) MILLER, J.C.; MILLER, J.N. Ellis Horwood Corp. Chichester. West Sussex. 1993.

A User-Friendly Guide to Multivariate Calibration and Classification NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. NIR Publications. Chichester. UK. 2004.

Handbook of Chemometrics and Qualimetrics (Part A – Estadística ; Part B – Calibratge)

MASSART, D.L.; WANDEGINSTE, B.G.M.; BUYDENS, L.M.C.; DeJONG, S.; LEWI, P.J.; SMEYERS-VERBEKE, J. Elsevier Science B.V. Amsterdam. The Netherlands. 1997 (A), 1998 (B).

Statistical Methods in Analytical Chemistry (2nd Ed.) MEIER, P.C.; ZÜND, R.E. John Wiley & Sons, Inc. Chichester. UK. 2000.