Universitat de Lleida

GUÍA DOCENTE

APLICACIONES DELANÁLISIS MULTIVARIANTE EN BIOTECNOLOGÍA

Coordinación: GATIUS CORTIELLA, FERNANDO

Año académico 2019-20

APLICACIONES DELANÁLISIS MULTIVARIANTE EN... 2019-20

Información general de la asignatura

Denominación	APLICACIONES DELANÁLISIS MULTIVARIANTE EN BIOTECNOLOGÍA
Código	101636
Semestre de impartición	2o Q(SEMESTRE) EVALUACIÓN CONTINUADA / 1R Q(SEMESTRE) EVALUACIÓN CONTINUADA
Carácter	

Grado/Máster	Curso	Carácter	Modalidad
Grado en Biotecnología	4	OPTATIVA	Presencial

Número de créditos de la asignatura (ECTS)	6
Tipo de actividad, créditos y grupos	

Tipo de actividad	PRALAB	PRAULA		TEORIA
Número de créditos	1	2	3	3
Número de grupos	2	2	2	2

Coordinación	GATIUS CORTIELLA, FERNANDO
Departamento/s	QUÍMICA
Distribución carga docente entre la clase presencial y el trabajo autónomo del estudiante	60 horas presenciales + 90 horas no presenciales por grupo (en cada grupo se imparte exactamente la misma docencia)
Información importante sobre tratamiento de datos	Consulte este enlace para obtener más información.
Idioma/es de impartición	Catalán
Distribución de créditos	Asignatura que se imparte ÍNTEGRAMENTE en dos bloques distintos de optativas: - GRUPO 1 (máximo 12 estudiantes): 6 créditos impartidos en el BLOQUE 1 de asignaturas optativas (las clases comienzan en SEPTIEMBRE) - GRUPO 2 (máximo 12 estudiantes): 6 créditos impartidos en el BLOQUE 3 de asignaturas optativas (las clases comienzan en FEBRERO) IMPORTANTE!!! En el momento de la matrícula el estudiante tiene que seleccionar uno de los dos grupos debido a que cada grupo está limitado a un NÚMERO MÁXIMO DE 12 ESTUDIANTES.

APLICACIONES DELANÁLISIS MULTIVARIANTE EN... 2019-20

Profesor/a (es/as)	Dirección electrónica\nprofesor/a (es/as)	Créditos impartidos por el profesorado	Horario de tutoría/lugar
GATIUS CORTIELLA, FERNANDO	fernando.gatius@udl.cat	12	

Información complementaria de la asignatura

Datos de contacto del profesor:

Ferran Gatius Cortiella (Departamento de Química. Universitat de Lleida)

Despacho A 0.07.2. Pasillo Planta Baja - Edificio A. Campus ETSEA. Av. Rovira Roure, 191. 25198 - Lleida
Teléfono núm. +34 973 00 36 54. E-mail: gatius@quimica.udl.cat

Objetivos académicos de la asignatura

El estudiante que supere la asignatura debe: (**Objetivos de conocimiento**)
• ENTENDER y UTILIZAR la terminología básica y los conceptos fundamentales que se abordan en la materia especificados en el programa de la misma.
• Saber utilizar los conceptos mostrados para INTERPRETAR resultados obtenidos del análisis de datos de distinta naturaleza, tanto en el ámbito univariante como multivariante.
• Conseguir una formación de carácter científico basada en el conocimiento de la metodología implicada en el tratamiento de datos, así como en la universalidad que confiere el carácter APLICADO de la asignatura a datos de diferentes tipos.
• Más generalmente, demostrar una preparación para el diseño de investigaciones sobre la base de los principios de relevancia, control y generalización.

El estudiante que supere la asignatura debe ser capaz de: (**Objetivos de capacidad**)
• PLANTEAR y DESARROLLAR el estudio y el tratamiento de diferentes tipos de datos en función de su carácter.
• Expresar correctamente el resultado de un estudio y realizar comparaciones de resultados numéricos con las herramientas y contrastes correspondientes.
• Plantear y ANALIZAR / INTERPRETAR los modelos implicados en el estudio descriptiva de un conjunto de datos en el ámbito multivariante para extraer las conclusiones pertinentes.
• ENTENDER y APLICAR las técnicas multivariantes de proyección para otros estudios aparte de los descriptivos, como son los implicados en el cálculo de relaciones entre diferentes variables y los procedimientos de clasificación de muestras.
• INTERPRETAR, DISCUTIR y EXTRAER CONCLUSIONES de estudios multivariantes de datos (en general), tanto si han sido realizados por uno mismo como por otra persona.

Competencias

Competencias generales. El graduado en Biotecnología debe:
Desarrollar aplicaciones y protocolos biotecnológicos para obtener productos de interés humano.
Trabajar en empresas biotecnológicas en la investigación, desarrollo o producción de bioproductos.
Diseñar proyectos biotecnológicos innovadores mediante identificación de aplicaciones, ideas de negocio, planes de trabajo e implementación de nuevas técnicas/equipos.
Transmitir información, ideas, problemas y soluciones a un público especializado o no.

Competencias específicas de la asignatura:
Entender y aplicar las técnicas de análisis de datos en diferentes situaciones.
Interpretar y extraer conclusiones de los estudios que impliquen tratamiento de datos

Contenidos fundamentales de la asignatura

En la mayor parte de estudios que se desarrollan se trabaja con muestras / datos a menudo definidos por muchas variables. En esta asignatura se presentan una serie de herramientas para poder tratar toda esta información de forma eficiente SIN la necesidad de conocimientos avanzados de matemáticas o estadística, se da una visión global del análisis y tratamiento de datos con estas herramientas y, sobre todo, de su APLICACIÓN para el estudio y extracción de conclusiones.
Para hacer esto se empieza (introducción) situando al alumno con el repaso inicial (muy breve) de los principales conceptos del tratamiento univariante de datos (que el alumno ya conoce de primer curso) que le permitirán llegar a la expresión correcta del resultado de un estudio ya hacer tratamientos básicos de los valores numéricos que definen una variable. En esta introducción se repasan conceptos estadísticos básicos que se conectan con la presentación y aplicación de las técnicas de proyección como herramientas para tratar datos de tipo multivariante. Estas herramientas se aplican a los estudios descriptivos de datos, a la relaciónentre variables (ámbito de la regresión) y en la clasificación de objetos atendiendo a grupos preestablecidos. Finalmente, la materia se completa con la aplicación de técnicas de diseño de experimentos tanto a nivel preliminar del estudio (screening) como para la optimización de diferentes procedimientos en base a las variables consideradas.
Hay que tener en cuenta que es una asignatura de análisis de datos APLICADA, en la que se empleará el software correspondiente para hacer los cálculos de los modelos implicados en un estudio, y en la que se evaluará la comprensión de los conceptos implicados en la misma mediante la iNTERPRETACIÓN de los fenómenos que muestran los diagramas generados en dichos modelos (interpretación que conducirá a la extracción de las conclusiones pertinentes). Son fundamentales las clases prácticas en las que se enseñará tanto a utilizar dicho software como interpretar los modelos / diagramas implicados en los diferentes estudios.

TEMARIO TEÓRICO

INTRODUCCIÓN: TRATAMIENTO DE DATOS EN ÁMBITO UNIVARIANTE
Errores en el análisis cuantitativo de resultados. Tipo. Repetibilidad / reproducibilidad. Estadísticos para las determinaciones con medidas repetidas. Distribuciones. Normalidad. Concepto de MUESTRA. La distribución de medias. Límites de confianza. Aplicaciones a la presentación de los resultados de un estudio. Contrastes de significación: comparación de la precisión y el resultado de 2 métodos. Comparación de más de 2 valores (estudios con factores – ANOVA). Otros estudios: comparación de frecuencias, detección de outliers, pruebas de normalidad, ... Conclusiones a partir de los contrastes de significación. Errores asociados.

DEFINICIÓN DEL ÁMBITO MULTIVARIANTE PARA EL ESTUDIO DE CASOS
Observaciones indirectas y correlaciones. Análisis Multivariante y Estadística Multivariante. Presentación de los datos multivariantes: la matriz de datos objeto de un estudio. Objetivos principales de las técnicas de análisis multivariante.

MÉTODOS DE PROYECCIÓN. ANÁLISIS POR COMPONENTES PRINCIPALES (PCA)
Estudio de datos definidos por variables correlacionadas. Planteamiento y problemas. El primer Componente Principal (PC): definición y obtención. Extensión del modelo a más PC's. Conceptos de scores y loadings (proyecciones / representación de muestras y variables). Varianza explicada y varianza residual de los datos en un modelo PCA. Residual y leverage de las muestras del estudio. Estudio de los outliers o muestras extrañas que se pueden encontrar en un estudio. Interpretación del modelo PCA y los diagramas asociados.

RELACIÓN ENTRE VARIABLES. LA REGRESIÓN MULTIVARIANTE
Estudio de la regresión en ámbito univariante. El ajuste por mínimos cuadrados. Extensión del estudio al ámbito multivariante. La regresión lineal múltiple (MRL). Definición del modelo MLR. Limitaciones y problemas de los modelos MLR. Primera solución a los problemas encontrados: la selección de variables. Solución definitiva a los problemas encontrados: la Regresión por Componentes Principales (PCR). Necesidad de los Componentes Principales en el ámbito de la regresión multivariante. Estudio del modelo: scores, loadings, varianzas, errores. Limitaciones / problemas de los modelos PCR. Solución a los problemas de PCR: la Regresión por Mínimo Cuadrados Parciales (PLS). Definición de los factores PLS. Tipo de modelo PLS. Estudio de los modelos: representaciones (scores, loadings) y errores asociados.

LA VALIDACIÓN DE LOS MODELOS MULTIVARIANTES
Calibración y validación de modelos: requerimientos de los conjuntos asociados. El concepto de "conjunto de test" en la validación de modelos. Otros métodos: la corrección por leverage y la validación cruzada. Estudio de los

errores. Medida del error experimental (estudios con repeticiones). Bondad de los modelos de regresión: RMSEP, RMSEC y otros parámetros.

MÉTODOS DE CLASIFICACIÓN MULTIVARIANTE DE DATOS
Clasificaciones supervisadas y no supervisadas. Aplicación de los métodos de proyección en la clasificación: el método SIMCA. Definición de los modelos que definen las clases. Definición de la metodología de clasificación. Parámetros. Análisis de los resultados (tablas y diagramas) del procedimiento de clasificación.

APLICACIONES DE LAS TÉCNICAS DE PROYECCIÓN AL DISEÑO DE EXPERIMENTOS
Diseños aplicados a estudios preliminares: diseños factoriales, Plackett-Burmann. Diseños de optimización: Diseño Central Compuesto (CCD), Box-Behnken.

ACTIVIDADES PRÁCTICAS

1. Ejercicios de repaso de estudios univariantes con hoja de cálculo: contrastes de hipótesis, análisis de varianza y regresión lineal.
2. Exploración de datos. Métodos de proyección. Análisis por Componentes Principales (PCA). Representación de la información (scores y loadings). Residuales y leverage.
3. Regresión por Componentes Principales (PCR). Parámetros de calibración. Distribuciones de muestras y variables en el modelo obtenido. Error del modelo. Interpretación. Definición del modelo: obtención de los coeficientes de regresión.
4. Regresión por Mínimos Cuadrados Parciales (PLS). Detección de muestras extrañas (outliers) y tratamiento de las mismas en la obtención del modelo. Interpretación del modelo encontrado: distribuciones de muestras y variables. Error del modelo y obtención de los coeficientes de regresión. Comparación PCR-PLS. Comparación de diferentes métodos de Validación de modelos.
5. Métodos de clasificación. Clasificación supervisada y no supervisada. Método SIMCA. Análisis de los resultados de la clasificación relativos a objetos, variables y modelos de las clases.
6. Diseño de experimentos. Diseños preliminares y de optimización. Modelos implicados y aplicación de los mismos.

Ejes metodológicos de la asignatura

Tipo de actividad	Descripción	Activ. presencial		Activ. no presencial		Evaluación	Tiempo total	
		Objetivos	Horas	Trabajo alumno	Horas	Horas	Horas	ECTS
Lección magistral	Clase magistral (Aula. Grupo grande)	Explicación de los principales conceptos	30	Estudio: Conocer, comprender y sintetizar conocimientos	**45**		**75**	**3**
Problemas y casos	Clase participativa (Aula. Grupo grande)	Resolución de problemas y casos (aplicación de conceptos explicados. Se plantea junto con el anterior)	-	Aprender a resolver problemas y casos (aplicar conocimientos a datos reales. Se plantea junto con el anterior)	**-**		**-**	**-**
Seminario	Clase participativa (Grupo medio)	Realización de actividades de discusión o aplicación	20	Resolver problemas y casos. Discutir.	**30**	(incluidas en las horas de actividad presencial del alumno)	**50**	**2**
Laborat.								

Aula de informática	Práctica en el aula de informática (Grupo medio)	Ejecución de la práctica: comprender conceptos, medir, aplicar, ...	10	Estudiar, entender y aplicar lo que se ha aprendido	**15**		**25**	**1**
Prácticas de campo								
Visitas								
Activ. dirigidas								
Otros								
Totales			**60**		**90**		**150**	**6**

Sistema de evaluación

Exámenes	**Prácticas**	**Análisis de casos y problemas**	**Otras actividades**
Ejercicios prácticos para aplicar las técnicas a datos reales	Discusión de ejemplos / ejercicios de aplicación de las técnicas a diferentes tipos de datos en diferentes situaciones.	Estudio en clase. Ejercicios.	

Tipo de actividad	**Actividad de Evaluación**		**Peso calificación**
	Procedimiento	**Número**	**(%)**
Lección magistral	Ejercicio de aplicación de las técnicas explicadas en datos concretos	1	**25**
Problemas y casos	Ejercicios de aplicación de las técnicas explicadas en datos reales adecuadas a cada tema del curso (preparación de los ejercicios de evaluación)	2	**45**
Seminario	Planteamiento de problemas, orientación sobre las técnicas necesarias para poder resolverlos, resolución comentada y debate o discusión sobre los puntos que los alumnos consideren necesarios	...	**30**
Laboratorio			
Aula de informática			
Prácticas de campo			
Visitas			
Actividades dirigidas	Orientaciones a los alumnos que lo consideren necesario para analizar los datos de algún trabajo que estén desarrollando y pueda requerir el uso de las técnicas explicadas en clase		
Otros			

Totales	**100**

Bibliografía y recursos de información

Bibliografía básica

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.

Bibliografía complementaria

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.