

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	36434
Nom	Estadística per a dades òmiques
Cicle	Grau
Crèdits ECTS	6.0
Curs acadèmic	2023 - 2024

Titulació/titulacions

Titulació	Centre	Curs	Període
1406 - Grau en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	3	Segon quadrimestre

Matèries

Titulació	Matèria	Caràcter
1406 - Grau en Ciència de Dades	11 - Salut	Obligatòria

Coordinació

Nom	Departament
AYALA GALLEGO, GUILLERMO	130 - Estadística i Investigació Operativa
DZUNKOVA, MARIA	194 - Genètica

RESUM

Es proposa una introducció a l'anàlisi estadística de dades òmics. Entre d'altres utilitzarem dades obtingudes amb microarrays i RNA-Seq. Estudiem els conceptes biològics bàsics previs. Es comenten les tècniques òmiques, el processat previ de la informació i els problemes estadístics. En particular, estudiem el problema de les comparacions múltiples. després vam mostrar seva aplicació a el problema de l'expressió diferencial. També considerem anàlisi de grups de gens. El software utilitzat és R i Bioconductor.

Les classes de teoria s'impartiran en castellà i les classes pràctiques i de laboratori segons consta en la fitxa de l'assignatura disponible en la web del grau.



CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

Un primer curs en Estadística així com un domini mitjà de l'Idlenguatge R.

COMPETÈNCIES

1406 - Grau en Ciència de Dades

- (CG01) Coneixement de matèries bàsiques i tecnologies, que li capacite per a l'aprenentatge de nous mètodes i tecnologies, així com que li dote d'una gran versatilitat per a adaptar-se a noves situacions.
- (CG03) Capacitat per a la realització de models, càlculs, informes, planificació de tasques i altres treballs anàlegs en l'àmbit específic de la Ciència de Dades.
- (CG07) Capacitat per prendre decisions de forma autònoma, elaborant de forma adequada i original, arguments raonats, podent obtenir així hipòtesis raonables i contrastables.
- (CT01) Ser capaçs d'accedir a eines d'informació (bibliogràfiques) i d'utilitzar-les apropiadament en el desenvolupament de les seves tasques quotidianes.
- (CT03) Habilitat per defensar el seu treball amb rigor i arguments, exposant-ho de forma adequada i precisa, recolzant-se en els mitjans necessaris.
- (CE12) Capacitat per dissenyar i engegar solucions basades en anàlisis de dades en l'àmbit de la medicina i dels negocis, tenint en compte els requisits específics d'aquest tipus de casos d'ús.
- (CE15) Capacitat per modelitzar i analitzar la incertesa en estudis basats en dades així com saber interpretar i contextualitzar els resultats obtinguts.
- (CB2) Que els estudiants sàpien aplicar els seus coneixements al seu treball o vocació d'una forma professional i posseïsquen les competències que solen demostrar-se per mitjà de l'elaboració i defensa d'arguments i la resolució de problemes dins de la seua àrea d'estudi.
- (CB5) Que els estudiants hagen desenvolupat aquelles habilitats d'aprenentatge necessàries per a emprendre estudis posteriors amb un alt grau d'autonomia.

RESULTATS DE L'APRENTATGE

Aprendre les característiques especials de les dades en biomedicina. Conèixer les dades òmics (proteòmiques, genòmics i metabòlics). Conèixer les aplicacions de la ciència de dades en biomedicina. Aplicar els mètodes coneguts en problemes de biomedicina.

En particular, realitzar una anàlisi completa amb microarrays i RNA-Seq amb R / Bioconductor amb especial interès en la utilització d'eines d'investigació reproducible com Rmarkdown.

(Avaluació de les competències CB5, CG7, CT3, CE15)



DESCRIPCIÓ DE CONTINGUTS

1. Part de biologia - generació de les dades òmiques

Introducció a les ciències òmiques.
Àcids nucleics i proteïnes.
Els genomes dels éssers vius.
Cromosomes, mutacions i l'herència.
Extracció d'àcids nucleics i la PCR.
Preparació dels àcids nucleics per a la seqüenciació.
Diferents plataformes de seqüenciació.
Tipus d'arxius en anàlisi genètica.
Expressió gènica i l'epigenètica.
La mutació i la variació poblacional.
Les bases de dades generals de seqüències.
El programa BLAST, anotació dels gens, mapatge de les seqüències.
El càncer: les mutacions gèniques i cromosòmiques.
Aplicacions de la seqüenciació d'ADN.

2. Introducció a les dades òmics

Introducció a dades òmics. R i Bioconductor. Anotació de genomes.

3. Dades

Microarrays. Obtenció i preprocessat de dades de microarrays. Mètode RMA.
GEO i Array Express.
RNA-Seq. Obtenció i preprocessat de dades de RNA-Seq. Alineament i normalització.

4. Expressió diferencial

Expressió diferencial marginal o gen a gen.
Microarrays. Selecció no específica. Comparacions múltiples. Mètode de Benjamini-Hochberg.
El q-valor. Mètode limma per a dades de microarrays.
RNA-Seq. Mètodes Edger i DESeq2.

5. Anàlisi de grups de gens

Expressió diferencial considerant grups de gens.
Grups de gens utilitzant GO i Kegg. Anàlisi de sobre representació. Anàlisi basat en enriquiment.

**VOLUM DE TREBALL**

ACTIVITAT	Hores	% Presencial
Classes de teoria	32,00	100
Pràctiques en laboratori	20,00	100
Pràctiques en aula	8,00	100
Estudi i treball autònom	60,00	0
Preparació d'activitats d'avaluació	24,00	0
Resolució de qüestionaris on-line	6,00	0
TOTAL	150,00	

METODOLOGIA DOCENT

Es plantegen classes magistrals així com classes pràctiques en aula d'informàtica. Es proporciona el material desenvolupat en un manual. Es proposaran qüestionaris setmanals.

(Avaluació de les competències CB5, CT1)

AVALUACIÓ

En la primera convocatòria, l'assignatura s'avaluarà atenent dos aspectes:

SE1: Prova objectiva, que es realitzarà en acabar la docència i que constarà de qüestions teoricopràctiques i problemes. Aquest apartat de l'avaluació comptarà un 70% de la nota final. D'aquest 70%, el 50% (sobre el total de l'assignatura) correspon a la part d'Estadística i el 20% restant (sobre el total de l'assignatura) correspon a la part de Biologia (avaluació de les competències CB4, CB5, CG1, CE9, CE15). Les classes de la part de biologia són obligatòries i l'assistència forma 10% de la nota de la part de biologia.

SE2: Realització de qüestionaris durant el curs avaluant, mitjançant preguntes de teoria i pràctica, el treball de l'estudiant. Aquesta part correspon completament a la part d'Estadística i suposa el 30% sobre la nota final. Aquesta part de l'avaluació no és recuperable (avaluació de les competències CB2, CB5, CG1,

CT2, CE9, CE15). El repartiment de les valoracions es farà de manera que el primer bloc de continguts biològics tindrà una valoració global del 20% i els restants blocs del 80%.

L'apartat SE2 no és recuperable en la segona convocatòria.

En la segona convocatòria es mantindrà el pes relatiu dels continguts de forma de manera que el primer bloc compte un 20% i els restants un 80%.

En qualsevol cas, el sistema d'avaluació es regirà pel que s'estableix en el Reglament d'Avaluació i Qualificació de la Universitat de València per a Graus i Màsters

(<https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion=inicio&idEdictoSeleccionado=5639>)



REFERÈNCIES

Bàsiques

- Bioinformàtica Estadística. Anàlisis estadístico de datos ómicos. Guillermo Ayala. <https://www.uv.es/ayala/docencia/tami/tami13.pdf>
- Pevsner J. (2015) Bioinformatics and Functional Genomics, 3rd Edition Wiley-Blackwell
- Genetics Home Reference (2020). Help Me Understand Genetics (<https://ghr.nlm.nih.gov/>)