Guia docent
200631 - ADO - Anàlisi de Dades Òmiques

Última modificació: 19/04/2022

Unitat responsable: Facultat de Matemàtiques i Estadística
Unitat que imparteix: 1004 - UB - Universitat de Barcelona.

Titulació: MÀSTER UNIVERSITARI EN ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (Pla 2013). (Assignatura optativa).

Curs: 2022 **Crèdits ECTS:** 5.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: DIEGO GARRIDO MARTÍN

Altres: Segon quadrimestre:
DIEGO GARRIDO MARTÍN - A
SANTIAGO RIOS AZUARA - A

CAPACITATS PRÈVIES

L'assignatura no pressuposa més coneixements previs que els habituals en un estudiant de Màster o Llicenciatura d'Estadística.
Malgrat això una bona predisposició cap a la biologia (concretament de Biologia molecular) i coneixement de programació i de llenguatge R son basics per tal d'obtenir el màxim profit de l'assignatura.
En un itinerari "ideal" aquesta assignatura vindria després d'una introducció a la bioinformàtica com la que conté el mateix programa d"estudis. Atès que ara per ara no es pot garantir aquesta situació ideal les dues assignatures són relativament independents de manera que, si bé resulta d'interès haver cursat "Fonaments de Bioinformàtica" per tenir certa familiaritat amb els problemes que es poden resoldre mitjançant les tècniques desenvolupades aquí, no es considera imprescindible.

REQUISITS

L'assignatura pressuposa uns nivells bàsic d'estadística com els que es poden assolir en el primer semestre del Master. Conve estar familiaritzat amb els conceptes de proves d'hipòtesis i significació estadística, anàlisi de la variància i tècniques bàsiques d'anàlisi multivariant: anàlisi de components principals i anàlisi de clusters.
Els conceptes necessaris per seguir el curs es poden trobar per exemple en el text "Applied Statistics for Bioinformatics using R" disponible a la web d'R (cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf)
o adicionalment a partir del llibre Data Analysis for the Life Sciences (http://rwdc2.com/files/rafa.pdf)

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:
5. CE-1. Capacitat per a dissenyar i gestionar la recollida d'informació, així com la codificació, manipulació, emmagatzematge i tractament d'aquesta informació.
6. CE-2. Capacitat per a dominar la terminologia pròpia d'algun àmbit en el que sigui necessària l'aplicació de models i mètodes estadístics o d'investigació operativa per a resoldre problemes reals.
7. CE-3. Capacitat per a formular, analitzar i validar models aplicables a problemes d'índole pràctica. Capacitat de seleccionar el mètode i/o tècnica estadística o d'investigació operativa més adequada per aplicar aquest model a cada situació o problema concret.
8. CE-5. Capacitat per a formular i resoldre problemes reals de presa de decissions als diferents àmbits d'aplicació sabent triar el mètode estadístic i l'algoritme d'optimització més adequat a cada ocasió.
9. CE-6. Capacitat per a fer servir el software més adequat per a realitzar els càlculs necessaris a la resolució d'un problema.
10. CE-9. Capacitat per a implementar algoritmes d'estadística i investigació operativa.

Transversals:
1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que en regeixen l'activitat; tenir capacitat per comprendre les regles laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici.

2. SOSTENIBILITAT I COMPROMÍS SOCIAL: Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; assolir habilitats per usar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.
3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

4. 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.

METODOLOGIES DOCENTS

L'enfocament de l'assignatura és teòrico-pràctic.
- Mitjançant casos guiats es presentaran els conceptes bàsics a l'aula
- Les tècniques empreades i la fonamentació tèorica associada es demostraran a l'aula
-Els alumnes complementaran els conceptes apresos mitjançant el seu treball personal
en activitats guiades i exercicis proposats.

La participació dels alumnes es dura a terme de tres formes
- Mitjançant la seva intervenció activa en les discussions plantejades (en línia) en forma de debats (al menys un per cada part del curs).
- Mitjançant la realització de petits exercicis proposats al llarg de l'assignatura amb periodicitat quinzenal.
- Mitjançant la realització i presentació de dos treballs pràctics (Per exemple una anàlisi de dades de microarrays i una segona anàlisi que pot ser de dades de NGS, o un altre tipus) .

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

La biologia mol·lecular i la biomedicina (i en paral·lel l'estadística) ha rebut una gran empenta en els darrers anys degut, entre altres raons, a la possibilitat de generar dades de forma massiva les més conegudes de les quals són les del genoma humà. Un cop han estat disponibles les seqüències dels genomes, i si fa no fa dels gens, la generació de dades no s'atura sinó que s'ha incrementat. Per exemple la tecnologia dels microarrays, amb gairebé deu anys de vida permet realitzar experiments on s'analitza de forma simultània l'expressió de tots els gens d'un individu amb finalitats com caracteritzar una certa situació patològica o de predir l'evolució d'un procés biològic. Tots aquests desenvolupaments han fet passar l'estadística al primer pla: sense ella no és possible accedir, manipular, depurar o analitzar aquestes grans quantitats d'informació.
L'objectiu d'aquesta assignatura és donar a conèixer els problemes que apareixen arrel de l'aparició de les tècniques de generació massiva de dades ("high throughput") i mostrar com s'hi aplica l'estadística (i la bioinformàtica) per afrontar-los. Aquesta aplicació es pot separar en dos aspectes
- D'una banda hi ha la utilització de mètodes estadístics convencionals a aquests nous problemes.
- D'altre banda apareix la necessitat de desenvolupar nous mètodes i noves eines per poder tractar aquestes noves dades.
Tots dos aspectes seran tractats en el curs..

Capacitats a adquirir
Les capacitats a adquirir al llarg d'aquest curs seran
- Coneixement dels diferents tipus de dades d'alt rendiment i les tècniques utilitzades per generar-les.
- Coneixement dels mètodes per tractar (recollir, preprocessar, analitzar, magatzemar) les dades d'alt rendiment, donant especial importància a la possibilitat de dur a terme un procés d'anàlisi completa: des de la generació fins a l'obtenció dels resultats.
- Coneixement dels mètodes i domini d'algunes de les eines existents per al seu tractment. Es donarà especial importància a la utilització de programari lliure i públic, i en especial al llenguatge R.

HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	80,0	64.00
Hores grup petit	15,0	12.00
Hores grup gran	30,0	24.00

Dedicació total: 125 h

CONTINGUTS

1. Introducció a la biologia mol·lecular, les òmiques i les tecnologies de generació de dades

Descripció:
1.1 Conceptes bàsics de biologia molecular
1.2 Mètodes d'obtenció de dades d'alt rendiment
1.2.1 Perspectiva general
1.2.2 Microarrays d'expressió génica
1.2.3 Altres tipus de dades (Ultraseqüenciació (NGS), Proteòmica, Metabolòmica...)

Dedicació: 6h
Grup gran/Teoria: 3h
Grup mitjà/Pràctiques: 3h

2. Anàlisi de dades de microarrays

Descripció:
2.1 Perspectiva general de l'anàlisi de dades de microarrays d'expressió
2.2 Lectura i control de qualitat de les imatges.
2.3 Preprocessat:Normalització i filtratge
2.4 Detecció de gens diferencialment expressats
2.4.1 Problemes estadístics que apareixen: potència i multiplicitat de proves.
2.5 Cerca de patrons de coexpressió mitjançant anàlisi de clusters
2.6 Diagnostics moleculars i mètodes de classificació.
2.6.1 Problemes estadístics que apareixen en l'elaboracio de predictors
2.7 L'ontologia génica i les seves aplicacions per a la interpretació biològica.

Dedicació: 20h
Grup gran/Teoria: 10h
Grup mitjà/Pràctiques: 10h

3. Anàlisi d'altres dades d'alt rendiment

Descripció:
3.1 Anàlisi de dades d'ultraseqüenciació: Visió general de les dades de NGS i de les tecnologies que les generen. Aplicacions
3.2. Control de qualitat de les dades de NGS. Preprocessat i correcció de problemes.
3.3 Anàlisi d'Expressió amb dades d'NGS
3.4. Altres aplicacions: cerca de variants en exomes i metagenòmica.

Dedicació: 14h
Grup gran/Teoria: 7h
Grup mitjà/Pràctiques: 7h

SISTEMA DE QUALIFICACIÓ

Es durà a terme avaluació contínua basada en la participació dels alumnes en cadascuna de les activitats descrites en l'apartat d'Organització. La valoració de cadascuna de les activitats serà:
· Participació en classe i en els debats: 10%
· Realització dels exercicis proposats a classe: 30%
· Realització de les proves d'avaluació continua proposades: 60%

BIBLIOGRAFIA

Bàsica:
- Draghici, S. Statistics and data analysis for microarrays using R and bioconductor [en línia]. 2nd ed. Chapman & Hall/CRC Mathematical & Computational Biology, 2012 [Consulta: 03/03/2021]. Disponible a: https://www.taylorfrancis.com/books/9780429130588.
- Tuimala, Jarno ; Laine, M. Minna. DNA microarray data analysis [en línia]. 2nd ed. CSC, the Finnish IT center for Science, 2005Disponible a: https://www.researchgate.net/publication/261680899_DNA_Microarray_Data_Analysis_second_edition. ISBN 9525520129.
- Gibson, G. ; Muse, S.V. A Primer of genome science. 3rd ed. 2012.
- Gentleman, R.; Carey, V.; Dudoit, S.; Irizarry, R.; Huber, W. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer, 2005.
- Irizarry, R.A; Love, M.I. Data Analysis for the Life Sciences [en línia]. 2015Disponible a: https://www.perlego.com/book/1573996/data-analysis-for-the-life-sciences-with-r-pdf.