Guía docente
200631 - ADO - Análisis de Datos Ómicos

Última modificación: 19/04/2022

Unidad responsable: Facultad de Matemáticas y Estadística
Unidad que imparte: 1004 - UB - Universitat de Barcelona.

Titulación: MÁSTER UNIVERSITARIO EN ESTADÍSTICA E INVESTIGACIÓN OPERATIVA (Plan 2013). (Asignatura optativa).

Curso: 2022 **Créditos ECTS:** 5.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: DIEGO GARRIDO MARTÍN

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

CAPACIDADES PREVIAS

La asignatura no presupone más conocimientos previos que los habituales en un estudiante de Master o Licenciatura de Estadística.
A pesar de eso una buena predisposición hacia la biología (específicamente biología molecular) y conocimiento de programación y de lenguaje R son básicos para obtener el máximo aprovechamiento de la asignatura.
En un itinerario "ideal" esta asignatura vendría después de una introducción a la bioinformática como la que contiene el mismo programa de estudios. Dado que hoy por hoy no se puede garantizar esta situación ideal las dos asignaturas son relativamente independientes de manera que, si bien resulta de interés haber cursado "Fundamentos de Bioinformática" para tener cierta familiaridad con los problemas que se pueden resolver mediante las técnicas desarrolladas aquí, no se considera imprescindible.

REQUISITOS

La asignatura presupone unos niveles básicos de estadística como los que se pueden alcanzar en el primer semestre del Master. Conviene estar familiarizado con los conceptos de pruebas de hipótesis y significación estadística, análisis de la varianza y técnicas básicas de análisis multivariante como el análisis de componentes principales y análisis de clusters.
Los conceptos necesarios para seguir el curso se pueden encontrar por ejemplo en el texto "Applied Statistics for Bioinformatics using R" disponible en la web de R (cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf) o adicionalmente a partir del libro Data Analysis for the Life Sciences (http://rwdc2.com/files/rafa.pdf)

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:
5. CE-1. Capacidad para diseñar y gestionar la recogida de información, así como la codificación, manipulación, almacenamiento y tratamiento de esta información.
6. CE-2. Capacidad para dominar la terminologia propia de algún ámbito en el que sea necesaria la aplicación de modelos y métodos estadísticos o de investigación operativa para resolver problemas reales.
7. CE-3. Capacidad para formular, analizar y validar modelos aplicables a problemas de índole práctica. Capacidad de seleccionar el método y/o técnica estadística o de investigación operativa más adecuado para aplicar dicho modelo a cada situación o problema concreto.
8. CE-5. Capacidad para formular y resolver problemas reales de toma de decisiones en los diferentes ámbitos de aplicación sabiendo elegir el método estadístico y el algoritmo de optimización más adecuado en cada ocasión.
9. CE-6. Capacidad para utilizar el software más adecuado para realizar los cálculos necesarios en la resolución de un problema.
10. CE-9. Capacidad para implementar algoritmos de estadística e investigación operativa.

Transversales:
1. EMPRENDIMIENTO E INNOVACIÓN: Conocer y entender la organización de una empresa y las ciencias que rigen su actividad; tener capacidad para entender las normas laborales y las relaciones entre la planificación, las estrategias industriales y comerciales, la calidad y el beneficio.

2. SOSTENIBILIDAD Y COMPROMISO SOCIAL: Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.
3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

4. USO SOLVENTE DE LOS RECURSOS DE INFORMACIÓN: Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad, y valorar de forma crítica los resultados de dicha gestión.

METODOLOGÍAS DOCENTES

El enfoque de la asignatura es teórico-práctico.
-Mediante la presentación de casos guiados se presentarán los conceptos básicos en el aula.
-Las técnicas utilizadas y su fundamentación teórica se demostrarán en el aula.
-El alumno complementará los conceptos aprendidos mediante su trabajo personal en actividades guiadas y ejercicios propuestos.

La participación de los alumnos se llevará a cabo de tres formas:
La participación de los alumnos se llevará a cabo de tres formas
- Mediante su intervención activa en las discusiones planteadas (en línea) en forma de debates (al menos uno por cada parte del curso).
- Mediante la realización de pequeños ejercicios propuestos a lo largo de la asignatura con periodicidad quincenal.
- Mediante la realización y presentación de dos trabajos prácticos (por ejemplo un análisis de datos de microarrays y un segundo análisis que puede ser de datos de NGS, u otro tipo).

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

La biología molecular y la biomedicina (y en paralelo la estadística) ha recibido un gran empuje en los últimos años debido, entre otras razones, a la posibilidad de generar datos de forma masiva, los más conocidos de los cuales son los del genoma humano. Una vez han estado disponibles las secuencias de los genomas, y más o menos de los genes, la generación de datos no se detiene sino que se ha incrementado. Por ejemplo la tecnología de los microarrays, con casi diez años de vida permite realizar experimentos donde se analiza de forma simultánea la expresión de todos los genes de un individuo con finalidades cómo caracterizar una cierta situación patológica o de predecir la evolución de un proceso biológico. Todos estos desarrollos han hecho pasar la estadística al primer plano: sin ella no es posible acceder, manipular, depurar o analizar estas grandes cantidades de información.

El objetivo de esta asignatura es dar a conocer los problemas que aparezcan a raíz de la aparición de las técnicas de generación masiva de datos ("high throughput") y mostrar cómo se aplica la estadística (y la bioinformática) para afrontarlos. Esta aplicación se puede separar en dos aspectos:
-Por una parte está la utilización de métodos estadísticos convencionales a estos nuevos problemas.
-Por otra parte aparece la necesidad de desarrollar nuevos métodos y nuevas herramientas para poder tratar estos nuevos datos.
Los dos aspectos serán tratados en el curso.

Capacidades a adquirir:
Las capacidades a adquirir a lo largo de este curso serán:
-Conocimiento de los diferentes tipos de datos de alto rendimiento y las técnicas utilizadas para generarlas.
-Conocimiento de los métodos para tratar (recoger, preprocesar, analizar, almacenar) los datos de alto rendimiento, dando especial importancia a la posibilidad de llevar a cabo un proceso de análisis completo: desde la generación hasta la obtención de los resultados.
-Conocimiento de los métodos y dominio de algunas de las herramientas existentes para su tratamiento. Se dará especial importancia a la utilización de software libre y público, y en especial al lenguaje R.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo pequeño	15,0	12.00
Horas grupo grande	30,0	24.00
Horas aprendizaje autónomo	80,0	64.00

Dedicación total: 125 h

CONTENIDOS

1. Introducción a la biología molecular, las ómicas y a las tecnologías de generación de datos

Descripción:
1.1 Conceptos básicos de biología molecular
1.2 Métodos de obtención de datos de alto rendimiento
1.2.1 Perspectiva general
1.2.2 Microarrays de expresión génica
1.2.3 Otros tipos de datos (Ultrasecuenciación (NGS), Proteómica, Metabolómica ...)

Dedicación: 6h
Grupo grande/Teoría: 3h
Grupo mediano/Prácticas: 3h

Análisis de datos de microarrays

Descripción:
2.1. Perspectiva general del análisis de datos de microarrays de expresión
2.2. Lectura y control de calidad de las imágenes
2.3. Preprocesado: Normalización y filtrado
2.4. Detección de genes diferencialmente expresados
2.4.1. Problemas estadísticos que aparecen: potencia y multiplicidad de pruebas.
2.5. Busca de patrones de coexpresión mediante análisis de clústers
2.6. Diagnósticos moleculares y métodos de clasificación
2.6.1. Problemas estadísticos que aparecen en la elaboración de predictores
2.7. La ontología génica y sus aplicaciones para la interpretación biológica
aplicacions per a la interpretació biològica.

Dedicación: 20h
Grupo grande/Teoría: 10h
Grupo mediano/Prácticas: 10h

Análisis de otros datos de alto rendimiento

Descripción:
3.1 Análisis de datos de ultrasecuenciación: Visión general de los datos de NGS y de las tecnologías que las generan. aplicaciones
3.2. Control de calidad de los datos de NGS. Preprocesado y corrección de problemas.
3.3 Análisis de Expresión con datos de NGS
3.4. Otras aplicaciones: búsqueda de variantes en exomas y metagenómica.

Dedicación: 14h
Grupo grande/Teoría: 7h
Grupo mediano/Prácticas: 7h

SISTEMA DE CALIFICACIÓN

Se llevará a cabo evaluación continua basada en la participación de los alumnos en cada una de las actividades descritas en el apartado de Organización. La valoración de cada una de las actividades será:

- Participación en los debates: 10%
- Realización de los ejercicios propuestos en clase: 30%
- Realización de las pruebas de evaluación continua propuestas: 60% (25% cada una)

BIBLIOGRAFÍA

Básica:

- Draghici, S. Statistics and data analysis for microarrays using R and bioconductor [en línea]. 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 línea]. 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 línea]. 2015Disponible a: https://www.perlego.com/book/1573996/data-analysis-for-the-life-sciences-with-r-pdf.

RECURSOS

Otros recursos:

A parte de los libros hay una gran cantidad de información libre y de calidad en Internet.
- El portal de Wentian Li: Un portal con todo tipo de información sobre análisis de datos de microarrays.
- StatWeb: Web con enlaces a programas, grupos, datos, etc.