

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

Código	42586
Nombre	Bioinformática estadística
Ciclo	Máster
Créditos ECTS	6.0
Curso académico	2021 - 2022

Titulación(es)

Titulación	Centro	Curso	Periodo
2116 - M.U. en Bioinformática 12-V.1	Escuela Técnica Superior de Ingeniería	1	Segundo cuatrimestre

Materias

Titulación	Materia	Caracter
2116 - M.U. en Bioinformática 12-V.1	3 - Bioinformática estadística	Obligatoria

Coordinación

Nombre	Departamento
AYALA GALLEGO, GUILLERMO	130 - Estadística e Investigación Operativa

RESUMEN

Este módulo se ocupa de datos biológicos de alto rendimiento. Bajo esta denominación genérica se engloban datos proporcionados por técnicas como los microarrays o la espectrometría de masas. Son datos donde la dimensión es muy superior al número de réplicas de que se dispone. Frente a unos cientos de observaciones tenemos dimensiones de varios miles. Algunos de los procedimientos estadísticos clásicos son utilizables mientras que otros han de modificarse para esta nueva situación. En este módulo se pretende abordar esta problemática.

Todos los tratamientos estadísticos se realización con R. En particular utilizaremos Bioconductor de un modo intensivo.



CONOCIMIENTOS PREVIOS

Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

Otros tipos de requisitos

Haber realizado un curso básico de Probabilidad y Estadística es muy conveniente. Sin embargo, el primer tema repasa de un modo somero lo básico de la Probabilidad y Estadística que utilizamos en la asignatura.

COMPETENCIAS

2116 - M.U. en Bioinformática 12-V.1

- Que los/las estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- Que los/las estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- Que los/las estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- Que los/las estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
- Ser capaces de acceder a la información necesaria (bases de datos, artículos científicos, etc.) y tener suficiente criterio para su interpretación y empleo.
- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Ser capaces de acceder a herramientas de información en otras áreas del conocimiento y utilizarlas apropiadamente.
- Ser capaces de valorar la necesidad de completar su formación científica, histórica, en lenguas, en informática, en literatura, en ética, social y humana en general, asistiendo a conferencias o cursos y/o realizando actividades complementarias, autoevaluando la aportación que la realización de estas actividades supone para su formación integral.
- Desarrollar la iniciativa personal y ser capaces de realizar una toma rápida y eficaz de decisiones en su labor profesional y/o investigadora.



- Trabajar en equipo con eficiencia en su labor profesional y/o investigadora y con personas de diferente procedencia.
- Aplicar las técnicas estadísticas básicas y las adaptadas al contexto del tratamiento estadístico computacional de muestras de origen experimental o clínico de alto rendimiento.

RESULTADOS DE APRENDIZAJE

- Conocimiento de la problemática estadística especial que se plantea con datos biológicos de alto rendimiento.
- Conocimiento de las técnicas básicas así como su modificación a este contexto.
- Manejo básico de R y manejo más avanzado de Bioconductor. Este software es libre y con una rapidez de desarrollo que quizás lo haga la opción más natural para mantenerse razonablemente actualizado en este mundo de un desarrollo tan rápido. Además al estar desarrollado en una gran parte por académicos la bibliografía y documentación es muy grande.

DESCRIPCIÓN DE CONTENIDOS

1. Conceptos básicos de Probabilidad y Estadística

Probabilidad condicional e independencia. Variable aleatoria. Valor esperado y varianza. Distribución de una variable aleatoria. Distribución conjunta y marginal. Algunas distribuciones importantes. Distribución muestral.

2. Control de calidad con datos biológicos de alto rendimiento

Correcció de fons i normalització en microarrays.

3. Contrastes de hipótesis en datos biológicos de alto rendimiento

Test de la t para dos muestras. Test de Kolmogorov-Smirnov. Comparaciones múltiples. Control de la tasa de falsos positivos. Contrastes de significación para conjuntos de variables.

4. Clasificación no supervisada

Método de las k-medias y k-medoids. Evaluación de la clasificación. Silueta. Clustering jerárquico

**5. Clasificación supervisada**

Arboles de clasificación. Evaluación de la clasificación.

6. Reducción de la dimensión

Análisis de componentes principales. Escalado multidimensional.

7. Diseño experimental en datos biológicos de alto rendimiento.

Diseños experimentales básicos con datos de alto rendimiento.

8. Técnicas de remuestreo

Aleatorización y bootstrap.

VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Clases de teoría	21,00	100
Prácticas en laboratorio	9,00	100
Preparación de actividades de evaluación	20,00	0
Preparación de clases de teoría	40,00	0
Preparación de clases prácticas y de problemas	50,00	0
Resolución de casos prácticos	10,00	0
TOTAL	150,00	

METODOLOGÍA DOCENTE

En las clases de teoría se plantearán problemas cuya resolución requiere la metodología correspondiente a cada tema. A continuación se introducirá la técnica estadística adecuada y se aplicará a la resolución de problemas utilizando software estadístico. Para la preparación de la asignatura el estudiante dispondrá de una colección de problemas, separados por temas, que tendrá que resolver por su cuenta.

Las sesiones de prácticas, en aula de informática y sincronizadas con la teoría, permitirán al estudiante aplicar estos procedimientos a la resolución de problemas.



EVALUACIÓN

La evaluación del aprendizaje de los conocimientos y competencias conseguidas por los estudiantes se hará de forma continuada a lo largo del curso. Para ello se propondrán 2 trabajos a lo largo del curso y un examen final. El examen final tendrá el valor de un 50 % mientras que cada trabajo contará un 25 %.

Será obligatoria la asistencia a todas las sesiones de teoría y práctica.

En las dos convocatorias:

SE1 Evaluación continua: mínimo 5 y máximo 15.

SE2 Actividades: mínimo 35 y máximo 70.

SE4 Exámenes 50.

REFERENCIAS

Básicas

- Referencia b1: Lee, J. K. (Ed.) Statistical Bioinformatics A Guide for Life and Biomedical Science Researchers Wiley-Blackwell, 2010
- Referencia b2: Wit, E. & McClure, J. Statistics for microarrays: design, analysis, and inference Wiley , 2004
- Referencia b3: Krijnen, W. P. Applied Statistics for Bioinformatics using R, 2009

Complementarias

- Referencia c1: Hahne, F.; Huber, W.; Gentleman, R. & Falcon, S. Bioconductor Case Studies Springer, 2008.

ADENDA COVID-19

Esta adenda solo se activará si la situación sanitaria lo requiere y previo acuerdo del Consejo de Gobierno