VNIVERSITAT Đ VALÈNCIA

COURSE DATA

Data Subject	
Code	42586
Name	Statistical bioinformatics
Cycle	Master's degree
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)			
Degree	**Center**	**Acad. year**	**Period**
2116 - M.U. en Bioinformática 12-V.1	School of Engineering	1	Second term

Subject-matter		
Degree	**Subject-matter**	**Character**
2116 - M.U. en Bioinformática 12-V.1	3 - Statistical bioinformatics	Obligatory

Coordination	
Name	**Department**
AYALA GALLEGO, GUILLERMO	130 - Statistics and Operational Research

SUMMARY

This module handles high-throughput biological data. Under this generic name are included data provided by techniques such as microarrays or mass spectrometry. Are data where the dimension is greater than the number of replicas available. Faced with a few hundred observations have dimensions of several thousand. Some of the classical statistical procedures are used while others have changed. This module aims to address this problem.

All statistical treatments were carried out using R. In particular a heavy use of Bioconductor

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

A basic course in probability and statistics is very convenient. However, the first chapter reviews the basics of probability and statistics that we use in the course.

OUTCOMES

2116 - M.U. en Bioinformática 12-V.1

- Students should apply acquired knowledge to solve problems in unfamiliar contexts within their field of study, including multidisciplinary scenarios.

- Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.

- Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.

- Students should demonstrate self-directed learning skills for continued academic growth.

- Be able to access the information required (databases, scientific articles, etc.) and to interpret and use it sensibly.

- Students should possess and understand foundational knowledge that enables original thinking and research in the field.

- Be able to access to information tools in other areas of knowledge and use them properly.

- To be able to assess the need to complete the scientific, historical, language, informatics, literature, ethics, social and human background in general, attending conferences, courses or doing complementary activities, self-assessing the contribution of these activities towards a comprehensive development.

- 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.

LEARNING OUTCOMES

- Knowledge of special statistical problem that arises with high-throughput biological data.
- Knowledge of basic techniques as well as modification to this context.
- Basic use of R and Bioconductor most advanced use. This software is free and with a speed of development that might do the most natural choice to keep reasonably current in this world of such rapid development. Also to be developed in a largely academic literature and documentation is very large.

DESCRIPTION OF CONTENTS

1. Basic Concepts of Probability and Statistics

Basic Concepts of Probability and Statistics. Conditional probability and independence. Random variable. Expected value and variance. Distribution of a random variable. Joint and marginal distribution. Some important distributions. Sampling distribution.

2. Quality control of high-throughput biological data

Background correction and normalization in microarrays.

3. Hypothesis testing in high-throughput biological data.

Hypothesis testing in high-throughput biological data. Two sample t-test. Kolmogorov-Smirnov test. Multiple comparisons. Controlling the rate of false positives. Tests of significance for sets of variables.

4. Clustering

Method of k-means and k-medoids. Evaluation of the classification. Silhouette.
Hierarchical clustering.

5. Supervised classification

Classification trees. Evaluation of the classification.

6. Dimension reduction.

Principal component analysis. Multidimensional scaling.

7. Design of Experiments in high-throughput biological experiments.

Basic experimental designs with high-throughput biological data.

8. Resampling techniques.

Randomization and bootstrap

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	21,00	100
Laboratory practices	9,00	100
Preparation of evaluation activities	20,00	0
Preparing lectures	40,00	0
Preparation of practical classes and problem	50,00	0
Resolution of case studies	10,00	0
TOTAL	150,00	

TEACHING METHODOLOGY

In the lectures will raise problems whose solution requires the methodology for each subject. The following will introduce the appropriate statistical technique and applied to problem solving using statistical software. For the preparation of the course the student will have a collection of problems, separated by subject, you will have to resolve on their own.
Practice sessions in a computer room and synchronized with the theory, allow the student to apply these procedures to solving problems.

EVALUATION

The assessment of learning knowledge and skills achieved by students will be continuously throughout the course. 2 This work will be proposed along the course and a final exam. The final exam will be worth 50% while each work will contribute 25%.
Will have to attend all sessions of theory and practice.

In the two calls:
SE1 Continuous assessment: minimum 5 and maximum 15.
SE2 Activities: minimum 35 and maximum 70.
SE4 Exams 50.

REFERENCES

Basic

- 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

Additional

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