BDA: Biomedical Data Analysis

Descriptive details concerning the subject:

· Name of the subject: Biomedical Data Analysis
· Code : BDI
· Type of subject: Optional
· Credits: 5
· ECTS: 5
· Total hours: 125
· Scheduling:
· Course: 1st Course
· Period: 1st term 1st 5 weeks
· Coordination: Arcadi Navarro
· Department: Ciències Experimentals i de la Salut
· Building:
· Lecture Times: 11:00-12:00

- Teaching staff:

· Language: English
· Lecturer: Arcadi Navarro
· Department: Ciències Experimentals i de la Salut

· Language: English
· Lecturer: Manuel Pastor
· Department: Ciències Experimentals i de la Salut

· Language: English
· Lecturer: Ferran Sanz
· Department: Ciències Experimentals i de la Salut

· Language: English
· Lecturer: Belén Lorente
· Department: Ciències Experimentals i de la Salut

· Language: English
· Lecturer: Laura Ines Furlong
· Department: Ciències Experimentals i de la Salut

· Language: English
· Lecturer: Michael Greenacre
· Department: Departament d'Economia i Empresa

Subject presentation:

This course focuses in how to use standard statistical methods (implemented in software packages) to analyze Biomedical data. After a general introduction on probability theory and statistical inference, an emphasis will be made on the most-common methods used to analyze multivariate data. Particular cases will be used as illustrative examples. The course will be centered on Frequentist Statistics, but overviews of Bayesian and Maximum Likelihood methods will be provided.

The course comprise 5 ECTS credits, implying 20/25 hours of plenary lectures, 10 hours of exercises and hands-on computer classes, 13/18 hours of reading and personal study, and 2 hours performing tests

The subject is based on the understanding of key methodological concepts and tools and on the application of common software used in labs around the world.

As this is a completely incremental subject, the student is advised of the need of strong interaction with the lecturers and the need of keeping the class material up to date.

The subject focuses on practical implementation of different types of tools for statistical inference. Thus, the methods covered are strongly based on the good understanding of basic principles of probability and programming. The course includes magistral lectures and hands-on exercises on the use of publicly available software packages.

The course will be evaluated by means of an exam, individual and based on short questions/answers, some problems and some text questions.

- Prerequisites in order to follow the itinerary:

Previous programming knowledge and notions of probability are required.

- Competences to be achieved in the subject

Instrumental:

1. Proficient reading/writing/listening scientific English related to the subject.
2. Knowledge of office software to do quality scientific presentations and

reports related to the subject.

Interpersonal:
1. Group work
2. Ability to solve by yourself a given problem

Systemic:
1. Analysis and synthesis abilities
2. Ability to search for information

Specific:
1. To understand the concept of probability.

2. To understand Bayes' Theorem.

3. To distinguish statistical description from inference.

4. To understand the concept of random variable.

5. To become familiar with central trend and dispersion measures.

6. To understand the concept of probability distribution.

7. To become familiar the most common kinds of distributions.

8. To master the graphical representation distributions and summary statistics.

9. To understand the concept confidence intervals and standard error.

10. To understand the concept of hypothesis testing.

11. To understand the concept of Type I and II errors.

12. To master the concept of ANOVA and its different designs.

13. To master the concept of contingency table and the relevant testing procedures.

14. To master the concept and procedures for Regression and Correlation Analysis.

15. To understand the concept of non-parametric tests.

16. To understand multivariate statistics and representation procedures.

17. To understand resampling methods.

18. To master the concept and procedures for Principal Components Analysis.

19. To understand the concept and procedures for Correspondence Analysis.

20. To understand the concepts of multiple regression and correlation.

21. To understand the concept and procedures for other non-linear regression methods.

22. To understand the concept and procedures for other non-linear regression methods.

23. To understand the concept of Generalized Linear Models

24. To understand the concept of Bayesian Statistics.

25. To understand the concept of Maximum Likelihood.

26. To become familiar with the multiple testing problem and its solutions.

27. To master the R software package as a tool for the implementation of the procedures under study.

- Learning aims:

To understand and apply algorithms and methods currently used in Biomedicine to perform statistical inference upon data.

- Evaluation:

General assessment criteria:

The evaluation will consist of a final exam at the end of the course, worth 80% and the evaluation of the practical exercises performed during the course.

Competence	Attainment	Assessment	Scheduling

Evaluation	indicator	procedure	
Instrumental			
1. Proficient reading/writing/listening scientific English related to the subject	Correct understanding of proposed exercises and correct final presentation	Implicit in the exam and exercises	End of term
2. Knowledge of office software to do quality scientific presentations and reports.	High quality presentation of results of exercises.	Implicit in exercises	Progressive
Interpersonal			
1. Group work	Ability to do team work both in programming and in preparing final presentation	Implicit in group-based exercises.	End of term
2. Ability to solve by yourself a given problem	Correct answer of set of pen and pencil exercises and final examination	Implicit in exercises and final exam	End of term
Systemic			
1. Analysis and synthesis abilities	Development of algorithms for proposed problems in molecular simulations	Implicit in exercises	Progressive
2. Ability to search for information	Complete final presentation	Implicit in exercises	Progressive
Specific			
1. To understand the concept of probability.		Exam	End of term
2. To understand Bayes' Theorem.		Exam	End of term
3. To distinguish		Exam	End of term

statistical description from inference.		
4. To understand the concept of random variable.	Exam	End of term
5. To become familiar with central trend and dispersion measures.	Exam	End of term
6. To understand the concept of probability distribution.	Exam	End of term
7. To become familiar the most common kinds of distributions.	Exam	End of term
8. To master the graphical representation distributions and summary statistics.	Exam	End of term
9. To understand the concept confidence intervals and standard error.	Exam	End of term
10. To understand the concept of hypothesis testing.	Exam	End of term
11. To understand the concept of Type I and II errors.	Exam	End of term
12. To master the concept of ANOVA and its different	Exam	End of term

designs.		
13. To master the concept of contingency table and the relevant testing procedures.	Exam	End of term
14. To master the concept and procedures for Regression and Correlation Analysis.	Exam	End of term
15. To understand the concept of non-parametric tests.	Exam	End of term
16. To understand multivariate statistics and representation procedures.	Exam	End of term
17. To understand resampling methods.	Exam	End of term
18. To master the concept and procedures for Principal Components Analysis.	Exam	End of term
19. To understand the concept and procedures for Correspondence Analysis.	Exam	End of term
20. To understand the concepts of multiple regression and correlation.	Exam	End of term
21. To understand	Exam	End of term

the concept and procedures for other non-linear regression methods.		
22. To understand the concept and procedures for other non-linear regression methods.	Exam	End of term
23. To understand the concept of Generalized Linear Models	Exam	End of term
24. To understand the concept of Bayesian Statistics.	Exam	End of term
25. To understand the concept of Maximum Likelihood.	Exam	End of term
26. To become familiar with the multiple testing problem and its solutions.	Exam	End of term
27. To master the R software package as a tool for the implementation of the procedures under study.	Exam	End of term

Contents

Contents 1: Block 1: Introduction to Statistics and Probability Theory..

Concepts	Procedures	Attitudes
Description and inference.		
Probability. Independence. Conditional probability. Bayes theorem	To be able to perform basic probability calculations.	
Uni- and bivariant statistics. Central trend and dispersion measures. Experimental errors. Random variables. Expected values. Estimators and estimation. Confidence intervals.	To be able to use the basic concepts descriptive statistics to perform calculations with them.	
	To be able to estimate different moments of common distributions.	
Common distributions. Central limit theorem.	To be able to compute confidence intervals.	
Descriptive statistics. Central tendency and spread measures. Graphical representations.	To be able to represent data in a number of ways.	

Contents 2: Block 2 Introduction to inferential Statistics.

Concepts	Procedures	Attitudes
	To be familiar with hypothersis testing.	
Hypothesis testing. Types I and II errors.	To be able to perform basic two-group tests (Student's-t)	
The t-Student test?. Two-tailed vs. one-tailed tests.		
Contingency tables. Overview of common statistical tests.	To be able to analyse basic cotingency tables.	
SPSS and R.	To become familiar with the SPSS and R packages.	

Contents 3: Block 3: Comparing groups. ANOVA, Non-parametic and resampling statistics.

Concepts	Procedures	Attitudes
One factor ANOVA. Advanced ANOVA. ANOVA assumptions.	To be able to perform ANOVA's and interpret their results.	

Non-parametric alternatives to ANOVA Assessing the required sample size.	To be able to select alternatives to ANOVA when appropriate.	
Resampling methods in statistics.	To be able to use a variety of resampling methods.	
Introduction to Correlation and Regression.	To be able to perform linear regressions. To be able to perform correlations.	

Contents 4: Block 4: Multivariate analysis..

Concepts	Procedures	Attitudes
	To be able to understand the multivariate nature of biomedical data.	
Multivariate statistics Variances/covariances matrix. Correlations matrix. Principal Components Analysis.	To be able to perform PCAs	
Correspondence Analysis.	To be able to perform correspondence analysis.	
Multiple Linear Regression and Correlation. Partial correlation.	To be able to perform multiple correlations and regressions.	
Logistic and other non-linear regressions.	To be able to perform non-linear regressions.	
Generalized Linear Models.	To be able to understand the linear nature of most tests.	

Contents 5: Block 5: Overview of Advanced statistics..

Conceptes	Procedures	Attitudes
Bayesian Statistics. Maximum Likelihood	To understand the basis of Bayesian statistics.	
Expression Microarrays, Whole	To use basic bayesian	

	analysis tools.
genome scans.	To perform basic maximum likelihood calculations.
The Multiple Tests problem.	To understand the statistical challenges under current biomedical data.
Text Mining.	To apply various algorithms for multiple testing correction.