MSc in Bioinformatics for Health Sciences

BDA. Biomedical Data Analysis

Syllabus Information

Academic Course: 2018/19

Academic Center: 804 - Official Postgraduate Programme in Biomedicine

Study: 8045 – Bioinformatics for Health Sciences - MSc

Subject: 30171 – BDA. Biomedical Data Analysis

Credits: 5.0

Course: 1st

Teaching languages: English

Teachers: Arcadi Navarro, Ferran Sanz, Laura I Furlong and Hafid Laayouni

Teaching Period: 1st term

Presentation

This course focuses in how to use standard statistical methods 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 focused on Frequentist Statistics, but basic overviews of Bayesian and Maximum Likelihood methods will be provided.

The course comprises 5 ECTS credits, implying 25 hours of plenary lectures, 15 hours of exercises and hands-on computer classes (please bring your own laptop for the hands-on sessions), 40-60 hours of reading and personal study, and 3 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 a good understanding of basic principles of probability and programming. The course includes lectures and hands-on exercises on the use of publicly available software packages.

The course will be evaluated by means of an individual exam and based on short questions/answers, some problems and some text questions. Additionally, a percentage of the final qualification will also be based on continuous evaluation focusing on homework assignments and hands-on practical assignments.

Associated skills

General:

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 and manage information from different sources

Specific competences:

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 of Generalized Linear Models
23. To understand the concept of Bayesian Statistics.
24. To understand the concept of Maximum Likelihood.

25. To become familiar with the multiple testing problem and its solutions.
26. To master the R software package as a tool for the implementation of the procedures under study.

Learning outcomes

Student successfully completing this course will be able to understand and apply algorithms and methods currently used in Biomedicine to perform statistical inference upon data.

Prerequisites

Previous basic programming skills and notions of probability are required.

Contents

Contents section 1: Introduction to Statistics and Probability Theory.

Concepts	Procedures
Description and inference	
Probability. Independece. 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 compute confidence intervals.
Common distributions. Central limit theorem. Descriptive statistics. Central tendency and spread measures. Graphical representations.	To be able to represent data in a number of ways. To be able to estimate different moments of common distributions.

Contents section 2: Introduction to Inferential Statistics

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

Contents section 3: Comparing groups. ANOVA, Linear Models

Concepts	Procedures
Advanced ANOVA. ANOVA assumptions.	To be able to perform more advanced ANOVA's design and interpret their results.
Non-parametric alternatives to ANOVA. Assessing the required sample size.	To be able to select alternatives to ANOVA when appropriate.
Introduction to Correlation and Regression.	To be able to perform linear regressions. To be able to perform correlations.
Multiple Linear Regression. Correlation coefficient. Partial correlation.	To be able to perform multiple regression analysis. Understand differences between simple and multiple regression. Understand the use of partial correlation

Contents section 4: More on Linear models, Multivariate Analysis and Resampling methods.

Concepts	Procedures
ANCOVA: Analysis of Covariance	
Logistic and other non-linear regressions.	To be able to perform nonlinear regressions.
Multivariate statistics. Variances/covariances matrix. Correlations matrix. Principal Components Analysis.	To be able to understand the multivariate nature of biomedical data. To be able to perform PCAs
Generalized Linear Models.	To be able to understand the linear nature of most tests.
Resampling methods in statistics.	To be able to use a variety of resampling methods.
Power Analysis and assessing the required sample size.	To understand Power analysis concept and use. To be able to assess the required sample size for basic statistical analysis.

Contents section 5: Overview of Advanced statistics.

Concepts	Procedures
Bayesian Statistics. Maximum Likelihood	To understand the basis of Bayesian statistics. To use basic Bayesian analysis tools. To perform basic maximum likelihood calculations.
None Parametric Statistics	Understand when and how to use non parametric statistics. Example of widely used non parametric tests
Expression Microarrays, Whole genome scans.	To understand the statistical challenges under current biomedical data.
The Multiple Tests problem.	To apply various algorithms for multiple testing correction.
Text Mining.	Recent examples on text mining methods

Teaching methods

The course is divided into 3 types of activities: those developed in class (plenary, seminar and practical sessions), directed work (assignments) and autonomous work.

Class sessions:

a) Plenary lessons: sessions where the teacher introduces and explains the course contents defined in the previous section. These sessions include definitions, introduction to the most relevant methodologies and examples of each model .

b) Seminar sessions: In the seminar sessions will take place correction of lists of exercises. Discussion of results and explanation. Is extremely important that student do the exercises before correction. Delivery of exercises (when asked for) is mandatory and will score 0 if exercises are not delivered on time without a justified cause.

c) Practical sessions

Hands-on sessions will take place in parallel with theory and seminar sessions. Goal is to acquire skills using appropriate software and to get familiar with the practice of statistical analysis. R software will be used in this part.

Evaluation

General assessment criteria:

The evaluation will consist of a final exam at the end of the course, worth 60%, the evaluation of the exercises performed during the course (25%) and the hands on exercises performed at the end of the course (15%).

Grading system:

Grades are between 0 and 10 and an overall 5 is needed to pass.

Bibliography and Information Resources

Basic bibliography

M.L. Samuel, J.A. Witmer, A. Shaffner. Statistics for the life Sciences.

Supplementary bibliography

Dalgaard. P. Introductory Statistics with R. Springer, 2004.

Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall, 1993.

Gentleman, R. Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health). Springer 2005.

Ewens, W. J., Grant, G. Statistical Methods in Bioinformatics: An Introduction (Statistics for Biology and Health). Springer 2006.

Lee P. M. Bayesian Statistics: An Introduction. Hodder Arnold; 2004.

Moore D.S., Notz W. I. Statistics: Concepts and Controversies. W. H. Freeman, 2005.

Sokal, R.R., Rohlf, J. F. Biometry. 3rd edition, W.H. Freeman, 1994

Rosner B. Fundamentals of biostatistics. Thomson Higher Education, cop 2006

Teaching resources

- https://www.otexts.org/book/biostat
- http://onlinestatbook.com/
- http://www.biostathandbook.com/