UAB
Universitat Autònoma de Barcelona

# Methods of Analysis in Health Sciences

Code: 104414
ECTS Credits: 6

2024/2025

| Degree | Type | Year |
|---|---|---|
| 2503740 Computational Mathematics and Data Analytics | OT | 4 |

## Contact

Name: Jose Barrera Gomez

Email: jose.barrera@uab.cat

## Teaching groups languages

You can view this information at the end of this document.

## Prerequisites

The student is supposed to be familiar with the binomial and the normal distributions, as well as with R.

## Objectives and Contextualisation

The main aims of the course are:

- Learn about the main types of study designs in the field of Epidemiology.

- Learn about the potential impact of both missing data and error measurement on the results of a statistical analysis.

- Learn about the main indicators to measure the presence of a disease or an exposure.

- Learn about the main indicators to measure the association between exposure and disease, specially in the case where both exposure and outcome are binary.

- Be able to identify the appropriate statistical tools for the assessment of the association between a given exposure (potential risk or protective factor) and a given health outcome, according to the characteristics of the study design, in the context of epidemiological studies.

- Learn about the design and implementation of an exact test according to the study design.

- Learn about the design and implementation of simulation studies related to concepts such as empirical power or sample size calculation.

- Be able to search scientific papers using PubMed efficiently.

- Get familiar with the reading of scientific papers.

- Be able to apply the concepts studied in the subject to solve exercises based in true epidemiological data.

- Improve the efficiency when programming in R to solve the practical tasks proposed during the course.

- Be able to write reproducible statistical reports using LaTeX and the R package knitr.

## Learning Outcomes

1. CM34 (Competence) Propose suitable statistical models for epidemiological studies.
2. CM34 (Competence) Propose suitable statistical models for epidemiological studies.
3. CM35 (Competence) Write technical reports that clearly express the results and conclusions of a bioscience study using vocabulary specific to the field of application.
4. KM29 (Knowledge) Recognise the most used statistical inference methods in bioinformatics.
5. KM30 (Knowledge) Identify the use of statistical knowledge in bioinformatics and in health science.
6. KM30 (Knowledge) Identify the use of statistical knowledge in bioinformatics and in health science.
7. KM31 (Knowledge) Identify the most used statistical inference methods in epidemiology studies.
8. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
9. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
10. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
11. SM38 (Skill) Use the most common databases in the field of health science.

## Content

*

1. Introduction to the contents. Introduction to reproducible research using the R package knitr.

2. PubMed: Searching scientific papers. Structure of a paper.

3. Classification of studies

(a) Topics in biostatistics

(b) Epidemiological studies

i. Notation

ii. Classification criteria

iii. Types of epidemiological study design: Randomised epidemiological trials, Cohort, Case-control, Case-crossover, Cross-sectional, Ecological

(c) Studies classification diagram

4. Classification of variables and related regression models

(a) According to the measure type

(b) According to the role in the study

(c) Types of explanatory variables

(d) Types of regression models according to the metric of the response variable

(e) Response variables of type time

5. Dealing with missing data

(a) Introduction

(b) Types of missing data

(c) Dealing with missing data

6. Example of statistical methods in Health Sciences: Integration of multiple imputation in cluster analysis

(a) Overview of cluster analysis

(b) Overview of multiple imputation

(c) Integration of multiple imputation in cluster analysis

(d) Software

7. Measures of disease presence

(a) Introduction

(b) Prevalence

i. Definition

ii. Estimation

iii. Comments

(c) Cumulative incidence

i. Definition

ii. Comments

(d) Incidence rate

i. Definition

ii. Comments

iii. Comparing two incidence rates

8. Measures of association between exposure and disease

(a) Introduction

(b) The relative risk

i. Definition

ii. Comments

(c) The odds ratio

i. The odds

ii. The odds ratio

iii. Comments

(d)Confidence intervals for OR and RR

(e) The attributable risk

i. Population attributable risk

ii. Exposure attributable risk

9. Causality, confusion and interaction

(a) Introduction

(b) Causality

(c) Confusion

(d) Interaction

10. Example of statistical methods in Health Sciences: Regression models with transformed variables. Interpretation and software

(a) Overview of the linear regression model

(b) Logarithm transformation in linear regression models. Why?

(c) Interpretation of results in the original scale of the variables

(d) Software

* Unless the requirements enforced by the health authorities demand a prioritization or reduction of these contents.

## Activities and Methodology

| Title | Hours | ECTS | Learning Outcomes |
|---|---|---|---|
| Type: Directed | | | |
| Theory sessions | 28 | 1.12 | |
| Type: Supervised | | | |
| Practice sessions | 28 | 1.12 | |
| Type: Autonomous | | | |
| Personal work | 94 | 3.76 | |

- Theory sessions: In these sessions, the different concepts of the subject as well as illustrative examples are introduced. Also, some exercises are proposed to be solved (usually requiring R usage). The methodology is based in the presentation and discussion of slides as well as the presentation of some additional materials (mainly news published in online media and scientific papers searched in PubMed).

- Practice sessions: In these sessions, several practical examples and exercises will be proposed. Activities related to R usage, PubMed search, papers reading and statistical analyses will be developed. Some of the proposed exercises will be mandatory.

- Seminars attendance: The Department of Mathematics and the UAB Statistical Service organize statistical seminars. The students and the teacher would attend some of them, depending on the topic and the schedule.

* The proposed teaching methodology may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

## Assessment

## Continous Assessment Activities

| Title | Weighting | Hours | ECTS | Learning Outcomes |
|---|---|---|---|---|
| Assignments in group | 30% | 0 | 0 | CM34, CM35, KM29, KM30, KM31, SM36, SM38 |
| Exam (or compensatory exam) | 50% | 0 | 0 | CM34, CM35, KM29, KM30, KM31, SM36, SM38 |
| Exercises in group | 20% | 0 | 0 | CM34, CM35, KM29, KM30, KM31, SM36, SM38 |

- Assiignments in grup during the course. Teacher could assess individual participation with oral questions.

- Exam (face-to-face).

- Optional compensatory exam (face-to-face). If the student attend the compensatory exam, its qualification will substitute the score in the previous, ordinary exam, regardless of the score obtained in bothe exams. To be able to participate in the compensatory exam, students must have been previously evaluated in a set of activities whose weight is equivalent to a minimum of two thirds of the total grade for the subject. You must also have obtained a minimum grade of 3.5 out of 10 in the average grade for the subject.

- The final scoring of the course out of 10, Q, will be:

Q = min{T, E}, if T is less than 4 or E is less than 3.5,

Q = (T + E) / 2, if T is greater than or equal to 4 and E is greater than or equal to 3.5,

where T and E are the scoring, out of 10, of the assignments and the exam, respectively.

- This subject does not offer the possibility of a single assessment (i.e. "evaluación única").

* Student's assessment may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.

## Bibliography

Basic: All concepts developed in the class sessions will be published at Moodle, including the slides that will be discussed in the theory sessions.

Further readings: Students interested in going further can explore the following items.

- Agresti, Alan. Categorical Data Analysis. Wiley, 3rd Edition, 2013.

- Breslow, N., N. Day. Statistical methods in cancer research. International Agency for Research on Cancer, 1980.

- Clayton D., Hills, M. Statistical models in epidemiology. Oxford University Press, 1993.

- Dalgaard, P. Introductory Statistics with R. Springer, 3rd Edition, 2002.

- dos Santos, I. Cancer epidemiology: principles and methods. International Agency for Research on Cancer, 1999.

- Gordis, L. Epidemiology. W.B. Saunders, 2004.

- Lachin, J.M. Biostatistical Methods: The Assessment of Relative Risks. Wiley, 2000.

- Motulsky, H.J. Intuitive Biostatistics. Oxford University Press, 1995.

- Rothman, K., Greenland, S. Modern epidemiology. Lippincott Williams & Wilkins, 1998.

- Rothman, K. Epidemiology: an introduction. Oxford University Press, 2002.

- Wassertheil-Smoller, S. Biostatistics and epidemiology: a primer for health and biomedical prefessionals. Springer, 3rd Edition, 2004.

## Software

- R
- RStudio
- LaTeX

## Language list

| Name | Group | Language | Semester | Turn |
|---|---|---|---|---|
| (PLAB) Practical laboratories | 1 | Catalan/Spanish | first semester | afternoon |
| (TE) Theory | 1 | Catalan/Spanish | first semester | afternoon |