

Degree programme	Type	Course
Applied Statistics	OP	4

Contact lecturer

Name : Jose Barrera Gomez

Email : jose.barrera@uab.cat

Prerequisites

Students attending this subject are supposed to have previously passed the subject "Statistics in Health Sciences".

Group languages

You can consult this information at the [end](#) of the document.

Objectives

The main aims of the course are:

- Learn the main characteristics of a epidemiological cross-sectional study.
- Learn how to design a health questionnaire.
- Learn how to create, clean and validate a dataset from the information contained in a health questionnaire.
- Learn to model the association between a health outcome of interest and a potential related exposure, in presence of potential confounding.
- Learn to model prevalences and rates using generalized linear models in a single population or in different subpopulations.
- Use R to handling and modelling cross-sectional data.
- Be able to write reproducible statistical reports using LaTeX and the R package knitr.

Learning outcomes

- CM14 (Propose the statistical model needed to analyse data sets belonging to real studies.) Propose the statistical model needed to analyse data sets belonging to real studies.
- KM17 (Recognise the statistical models for the analysis of data with different structures and complexities that frequently appear in different fields of application.) Recognise the statistical models for the analysis of data with different structures and complexities that frequently appear in different fields of application.
- KM18 (Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.) Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.
- SM16 (Select appropriate sources of information for the statistical work.) Select appropriate sources of

information for the statistical work.

- SM17 (Discuss scientific articles in which the analysis of a study of the different areas of application is considered.) Discuss scientific articles in which the analysis of a study of the different areas of application is considered.
- SM18 (Refine the information available for subsequent statistical processing.) Refine the information available for subsequent statistical processing.
- SM19 (Analyse complex data, whether this is due to their characteristics or their size.) Analyse complex data, whether this is due to their characteristics or their size.

Contents

(*)

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

2. Cross-sectional data

(a) Cross-sectional data

(b) Information sources: Reported information, Measured information

(c) Aspects to consider during the design of a health survey

(d) The codebook

3. Population based studies: cross-sectional studies

(a) Characteristics

(b) Advantages

(c) Disadvantages

(d) Comparison with other epidemiological study designs

4. Measuring the disease presence in cross-sectional studies: the prevalence

5. Binary exposure and disease: the 2×2 contingency table

(a) Tests for independence between exposure and disease: Asymptotic approximation: the chi-square test, Fisher test: drawbacks, Design and implementation of an exact test under cross-sectional design

6. GLM overview.

(a) Model specification

(b) Maximum likelihood estimation of the parameters of the model

(c) Hypothesis tests for the parameters of the model: Wald test and likelihood ratio test

(d) Interpretation of the parameters of the model

(e) Dealing with confounders

(f) Considering interactions

(g) Validation

7. Modeling prevalences with the GLM.

(a) Modeling OR with logistic regression

(b) Modeling PR with log-binomial regression

(c) Modeling PD with lineal regression

(d) Goodness of fit

8. Modeling counts and rates with the GLM

(a) Poisson regression

(b) Binomial-negative regression

(c) Models for excess of zeros

9. Introduction to regression models for polytomous outcomes.

10. The Generalized Linear Mixed Model for modelling prevalences and rates in clustered data.

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

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Personal work	108	4.32	
Theory sessions	14	0.56	
Practice sessions	28	1.12	

(*)

- 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.

Use of Artificial Intelligence (AI) Technologies:

For this course, a restricted use of Artificial Intelligence (AI) technologies is permitted. Specifically, the use of AI is allowed exclusively for the following tasks:

- Literature or information searching
- Linguistic correction and/or translation of texts
- Assistance with LaTeX code
- Reviewing R code previously created by the student without the use of AI
- Tasks for which the professor explicitly permits, recommends, or mandates the use of AI

For each and every gradable written assignment submitted, the student (or group of students) must clearly identify which parts were generated using AI and, for each of these parts, must specify:

- Name and version of the AI tool
- Prompt (the question or questions asked to the AI)
- Critical reflection on how the use of the AI tool influenced both the process and the final outcome of the activity

Lack of transparency regarding the use of AI in these gradable activities will be considered a breach of academic honesty and may lead to a partial or total grade penalty for the activity, or more severe sanctions in serious cases.

In general, committing any irregularity in any of the assessment activities (academic fraud, plagiarism, or misuse of AI, unless such use is expressly authorized in the course guide) that could lead to a significant variation in the grade will result in a score of 0 for that specific activity. If the course guide states that achieving a minimum grade in that specific assessment activity is an essential requirement to pass the course, or if multiple irregularities occur across the course's assessment activities, the final grade for the course will be 0. Independent of this, disciplinary proceedings may be initiated against any student who commits such irregularities.

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

Continuous assessment activities

Title	Weight	Hours	ECTS	Learning outcomes
Exercises in group	20%	0	0	CM14, KM17, KM18, SM16, SM17, SM18, SM19
Assignments in group	30%	0	0	CM14, KM17, KM18, SM16, SM17, SM18, SM19

Exam (or compensatory exam)	50%	0	0	CM14, KM17, KM18, SM19
-----------------------------	-----	---	---	------------------------

(*)

- Assignments in group during the course. Teacher can formulate oral questions in order to assess individual contributions.
- 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 both exams.
- 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.

Use of Artificial Intelligence (AI) Technologies:

For this course, a restricted use of Artificial Intelligence (AI) technologies is permitted. Specifically, the use of AI is allowed exclusively for the following tasks:

- Literature or information searching
- Linguistic correction and/or translation of texts
- Assistance with LaTeX code
- Reviewing R code previously created by the student without the use of AI
- Tasks for which the professor explicitly permits, recommends, or mandates the use of AI

For each and every gradable written assignment submitted, the student (or group of students) must clearly identify which parts were generated using AI and, for each of these parts, must specify:

- Name and version of the AI tool
- Prompt (the question or questions asked to the AI)
- Critical reflection on how the use of the AI tool influenced both the process and the final outcome of the activity

Lack of transparency regarding the use of AI in these gradable activities will be considered a breach of academic honesty and may lead to a partial or total grade penalty for the activity, or more severe sanctions in serious cases.

In general, committing any irregularity in any of the assessment activities (academic fraud, plagiarism, or

misuse of AI, unless such use is expressly authorized in the course guide) that could lead to a significant variation in the grade will result in a score of 0 for that specific activity. If the course guide states that achieving a minimum grade in that specific assessment activity is an essential requirement to pass the course, or if multiple irregularities occur across the course's assessment activities, the final grade for the course will be 0. Independent of this, disciplinary proceedings may be initiated against any student who commits such irregularities.

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.
- Christensen, R. Log-Linear Models and Logistic Regression. Springer, 2nd Edition, 1990.
- 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.
- Hosmer, D.W., Lemeshow, S. Applied Logistic Regression. Wiley, 2nd Edition, 2000.
- Kleinbaum, D.G. y Klein, M. Logistic Regression. A Self-Learning Text. Springer, 2002.
- Lachin, J.M. Biostatistical Methods: The Assessment of Relative Risks. Wiley, 2000.
- Motulsky, H.J. Intuitive Biostatistics. Oxford University Press, 1995.
- McCullagh, P., Nelder, J.A. Generalized Linear Models. Chapman and Hall, 1983.
- 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 professionals. Springer, 3rd Edition, 2004

Software

- R
- RStudio
- LaTeX

Course groups and languages

The information provided is provisional until November 30. After this date, you will be able to consult the language of each group through this [link](#). To access the information, you will need to enter the course

CODE

Type of teaching	Group	Language	Semester	Shift
(TE) Theory	1	Catalan/Spanish	first semester	morning-mixed
(PLAB) Practical laboratories	1	Catalan/Spanish	first semester	morning-mixed