UAB
Universitat Autònoma de Barcelona

2023/2024

Cross-Sectional Data Analysis

Code: 104878
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	OT	4	1

Contact

Name: Jose Barrera Gomez

Email: jose.barrera@uab.cat

Teaching groups languages

You can check it through this link. To consult the language you will need to enter the CODE of the subject. Please note that this information is provisional until 30 November 2023.

Prerequisites

Students attending this subject are supposed to having previously attended the subject "Statistics in Health Sciences".

Objectives and Contextualisation

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.

Competences

- Analyse data using statistical methods and techniques, working with data of different types.
- Correctly use a wide range of statistical software and programming languages, choosing the best one for each analysis, and adapting it to new necessities.
- Critically and rigorously assess one's own work as well as that of others.
- Formulate statistical hypotheses and develop strategies to confirm or refute them.
- Identify the usefulness of statistics in different areas of knowledge and apply it correctly in order to obtain relevant conclusions.
- Interpret results, draw conclusions and write up technical reports in the field of statistics.
- Make efficient use of the literature and digital resources to obtain information.
- Select and apply the most suitable procedures for statistical modelling and analysis of complex data.
- Select statistical models or techniques for application in studies and real-world problems, and know the tools for validating them.
- Select the sources and techniques for acquiring and managing data for statistical processing purposes.
- Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
- Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
- Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Use quality criteria to critically assess the work done.
- Work cooperatively in a multidisciplinary context, respecting the roles of the different members of the team.

Learning Outcomes

1. Analyse data corresponding to epidemiological studies or clinical trials.
2. Carry out the most suitable sampling for epidemiological studies.
3. Critically assess the work done on the basis of quality criteria.
4. Design and conduct hypothesis tests in the different fields of application studied.
5. Draw conclusions that are consistent with the experimental context specific to the discipline, based on the results obtained.
6. Draw up technical reports that clearly express the results and conclusions of the study using vocabulary specific to the field of application.
7. Identify the most commonly statistical inference techniques used in epidemiology studies.
8. Interpret statistical results in applied contexts.
9. Justify the choice of method for each particular application context.
10. Make effective use of references and electronic resources to obtain information.
11. Reappraise one's own ideas and those of others through rigorous, critical reflection.
12. Recognize the advantages and drawbacks of the different statistical methodologies when studying data from a variety of disciplines.
13. Recognize the importance of the statistical methods studied within each particular application.
14. Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
15. Students must be capable of collecting and interpreting relevant data (usually within their area of study) in order to make statements that reflect social, scientific or ethical relevant issues.
16. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
17. Use different programmes, both open-source and commercial, associated with the different applied branches.
18. Work cooperatively in a multidisciplinary context, accepting and respecting the roles of the different team members.

Content

(*)

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.

Methodology

(*)

- 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.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Theory sessions	14	0.56	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 13, 18, 17, 10
Type: Supervised			
Practice sessions	28	1.12	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 13, 18, 17, 10
Type: Autonomous			
Personal work	108	4.32	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 13, 18, 17, 10

Assessment

(*)

- 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.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Assignments in group	30%	0	0	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 12, 13, 18, 17, 10
Exam (or compensatory exam)	50%	0	0	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 12, 13, 18, 17, 10
Exercises in group	20%	0	0	1, 11, 3, 4, 6, 5, 2, 7, 8, 9, 16, 14, 15, 12, 13, 18, 17, 10

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 prefessionals. Springer, 3rd Edition, 2004

Software

- R

- RStudio

- LaTeX