Universitat Autònoma de Barcelona

2023/2024

Longitudinal Data Analysis

Code: 104879
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	OT	4	2

Contact

Name: Juan Ramon Gonzalez Ruiz

Email: juanramon.gonzalez@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

- It is recommended, but not essential because a class will be held to homogenize the level of the students, to know the generalized linear models and the Cox model for the survival analysis.

- The subject on observational studies contains a complementary introduction to the topic that will deal with the analysis of the temporal evolution of incidence and mortality rates in which the concept of rate will be introduced.

- It is advisable but not essential to know the statistical package R.

Objectives and Contextualisation

The main objectives are:

- To know the statistical models for the analysis of longitudinal data (information obtained from measurements made over time) that often appear in health sciences (biology, medicine, pharmacology, toxicology, chemistry and / or engineering))

- To know the statistical models to analyze the temporal evolution of the incidence and mortality rates of a disease to detect temporary changes and why they are due

- To know the statistical models to analyze the time until the occurrence of an event of interest that appears recurrent (tumor relapse, migraine, heart attacks, ...) taking into account the effect of covariates, the effect of the intervention and / or the effect to observe several events beforehand

- To know the statistical models to analyze data obtained from repeated measurements over time using linear models (rehospitalizations, relapse of a disease, ...)

- To know the statistical models to analyze data obtained from repeated measurements over time using non-linear models (tumor growth in rats, evolution of the weight of children after birth, ...)

- To be able to read critically a scientific article that considers the analysis of a study in which information collected over time is available.

- Be able to identify the statistical model necessary to analyze a set of data that will be presented in practical exercises and that belong to real studies.

- Know how to perform all these analyzes using R using the appropriate libraries.

Competences

- 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.
- 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. Critically assess the work done on the basis of quality criteria.
2. Design and conduct hypothesis tests in the different fields of application studied.
3. Draw conclusions that are consistent with the experimental context specific to the discipline, based on the results obtained.
4. Draw up technical reports that clearly express the results and conclusions of the study using vocabulary specific to the field of application.
5. Interpret statistical results in applied contexts.
6. Justify the choice of method for each particular application context.
7. Make effective use of references and electronic resources to obtain information.
8. Reappraise one's own ideas and those of others through rigorous, critical reflection.
9. Recognize the importance of the statistical methods studied within each particular application.
10. 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.
11. 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.
12. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
13. Use different programmes, both open-source and commercial, associated with the different applied branches.
14. Work cooperatively in a multidisciplinary context, accepting and respecting the roles of the different team members.

Content

These are the proposed contents*:

1. Introduction to the course

2. Linear models for continuous longitudinal data: data structure, visualization and models with normal response

3. Longitudinal models for continuous dates: mixed models

4. Longitudinal data models with non-normal response

5. Picewise and segmented regression

6. Survival analysis with time-dependent variables

7. Analysis of survival with recurrent events

    7.1. Extension of the Cox model

    7.2. frailty models

    7.3. General model for recurring events

    7.4. Models with terminal event

8. Sample size calculation for longitudinal data

9. Joint Models (Survival + Longitudinal markers)

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

Methodology

Theoretical sessions:
In these sessions the main concepts of each topic will be presented, as well as the analysis of data with real examples in which the necessary code of R will be displayed to carry out this task. The slides (created with R Markdown - which guarantees reproducibility of the results) will include the theoretical concepts, the data analysis and the interpretation of the
results and the conclusions drawn from them.

Practical sessions:
In these sessions some guided exercises will be proposed that the student will have to solve individually. Each student will have an individual data set for the same problem. A random database for the same real problem will be generated independently for each student. With this methodology, the student is expected to investigate and learn how to analyze a set of
real data, regardless of whether he asks a classmate how to carry it out, since each student will have to analyze their data and obtain their conclusions from their own results.

Assistance to seminars:
Exceptionally, if there are certain time matches and if the students see it feasible to do it, the students, together

with the teacher, will attend a seminar organized by the UAB
Applied Statistics Service or by another research center close to the College. This assistance will not be compulsory, but it will be highly recommended by the teacher, since the students could see how the methodology they are learning is used in real studies and they could see how the work of a statistician has a crucial implication in the completion of said investigations.

Individualized work
The solutions of all the practices that are carried out in person must be delivered to the teacher. As already mentioned, these practices will be carried out individually, as each student will have a personalized database on the same problem to be addressed. In addition to these face-to-face practices, the studentwill have to solve four other practices at home and deliver the numerical solution, as well as the R code used to obtain said results. Both the classroom practices and those carried out at home will be part of the continuous evaluation. The student will have access to all the teaching material from a teaching web in which there will also be a forum in which to raise questions that could ideally be resolved by their peers and that the teacher will supervise and / or solve, if necessary.

PS: 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			
Theoretical sessions	150	6	8, 1, 2, 4, 3, 5, 6, 12, 10, 11, 9, 14, 13, 7

Assessment

Self-evaluation:
After each theoretical session, a series of general questions will be proposed to evaluate if the student has learned the basic concepts of the topic treated during that session.

Delivery of practices:
During the course the student will have to solve several practices in which the teacher will be able to guide since they will be done in person. The numerical solution must be sent
as well as the R code used to solve them.

Final Exam:
Students will take a face-to-face test to assess whether they have acquired the minimum theoretical and practical concepts about the subject. This test will contain conceptual
questions about the models treated in class and R outputs similar to those obtained in the analyzes that have been made during the course, on which questions will be asked to
interpret the results.

Final mark will be:

0,5*NE + 0.5*NPA

where NE stands for the exam and NPA for practical and self-evaluation marks, respectively. It is mandatory to have NE>=5 to pass the topic.

Single evaluation (optional):

A comprehensive exam (4 hours) will be conducted to assess the knowledge and skills acquired throughout the course. This exam will be designed to evaluate the student's ability to apply the statistical analyses learned and their understanding of theoretical concepts.

The exam will consist of two main parts: statistical analysis and theoretical questions. In the statistical analysis section, relevant data will be provided, requiring the student to apply the statistical techniques and tools learned during the course. The following steps are expected from the student:

1. Problem identification: The student should understand the nature of the data and the analysis objectives.

2. Selection and application of techniques: The student will use the acquired knowledge to selectand apply appropriate statistical techniques to analyze the data. This may include determining measures of central tendency, dispersion, correlation, regression, hypothesis testing, among others.

3. Interpretation of results: Once the analyses are performed, the student should interpret the results accurately, explaining their significance in the context of the given problem.

The second part of the exam will consist of theoretical questions that require written responses. These questions will be related to fundamental statistical concepts, their applicability in different situations, and their importance in decision-making. The student should demonstrate a clear understanding of the concepts and the ability to explain them coherently.

The evaluation of this exam will consider several criteria:

1. Accuracy and correctness in analyses: The student's ability to perform statistical analyses accurately and correctly will be evaluated, including selecting appropriate techniques and using the correct procedures.

2. Interpretation of results: The student's capacity to interpret and explain the results obtained from the statistical analyses will be assessed.

3. Completeness of theoretical responses: The student's ability to provide clear and comprehensive answers to the theoretical questions, demonstrating mastery of the concepts and their application, will be considered.

4. Organization and clarity of presentation: The overall organization of the exam, clarity of written responses, and quality of statistical result presentation will be taken into account.

PS: 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
Individual tasks	50%	0	0	8, 1, 2, 4, 5, 6, 12, 10, 11, 14, 13, 7
Test	50%	0	0	3, 5, 6, 9

Bibliography

González JR, Llorca F, Moreno V. Algunos aspectos metodológicos sobre los modelos edad-periodo-cohorte. Aplicación a las tasas de mortalidad por cáncer. Gaceta Sanitaria, 2002;16:267-273

Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Statistics in Medicien, 2000;19:335-51

Fernandez E, Gonzalez JR, JM Borras, et al. Recent decline in cancer mortality in Catalonia (Spain). A Joint point regression analysis. European Journal of Cancer, 2001:37:2222-2228.

Gonzalez JR, Peña E, Slate E. Modelling intervention effects alter cancer relapses. Statistics in Medicine, 2005:24:3959-1975

V Rondeau, Gonzalez JR. Frailtypack: a computer program for the análisis of correlated failure time data using penalized likelihood estimation. Computer Methods and Programs in Biomedicine, 2005;80:154-164.

González JR, Peña E. Estimación no paramétrica de la función de supervivencia para datos con eventos recurrentes. Revista Española de Salud Pública, 2004;78:211-220

Gonzalez JR. Modelling recurrent event data with application to cancer research. VDM Verlag, Saarbrken, Germany, 2009 (pdf del libro accesible en la web de la asignatura)

Therneau T and Grambsch P. Modeling Survival Data: Extending the Cox Model. Springer-Verlag, New York, 2000.

Software

Theroy and practical lectures will be in R.