Universitat Autònoma de Barcelona

2019/2020

Longitudinal Data Analysis

Code: 104879
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	OT	4	0

Contact

Name: Juan Ramón González Ruíz
Email: JuanRamon.Gonzalez@uab.cat

Use of Languages

Principal working language: spanish (spa)
Some groups entirely in English: No
Some groups entirely in Catalan: No
Some groups entirely in Spanish: No

Prerequisites

- It is recommended, but not essential because a class will be held to homogenize the level of the students, to kn

- The subject on observational studies contains a complementary introduction to the topic that will deal with the a

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 (infor

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 know the statistical models to analyze the time until the occurrence of an event of interest that appears recurr

taking into account the effect of covariates, the effect of the intervention and / or the effect to observe several ev

-
To know the statistical models to analyze data obtained from repeated measurements over time using linear mod

-
To know the statistical models to analyze data obtained from repeated measurements over time using non-linear

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 collecte
- Be able to identify the statistical model necessary to analyze a set of da

- 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

1. Introduction to the course
    1.1 Introduction to R markdown: creation of automated and reproducibl
    1.2 Tidyverse
2. Analysis of incidence and mortality rates
    a. Introduction
    b. Rate definition
    c. Calculation of standardized rates
    d. Analysis of temporal trends
    e. Regression models 'jointpoint'
    f. Age-period-cohort models
3. Survival analysis for data with recurrent events
    a. Introduction
    b. Non-parametric models
        i. Model Peña-Strawderman-Hollander
        ii. Model of Chan-Wang
        iii. Fragility model ('frailty model')
    c. Semi-parametric models
        i. Conditional model (Prentice-William-Peterson)
        ii. Marginal model (Wei-Lin-Weidsfeld)
        iii. Fragility model
        iv. General model (Peña-Hollander)
        v. Cancer model (González-Peña-Slate)
    d. Model with terminal event
        i. Estimation through penalized credibility
4. Analysis of longitudinal data by linear models
    a. Introduction
    b. Designs with repeated measures
    c. Repeated measures ANOVA
    d. MANOVA
    e. Mixed linear model.

f. Diagnosis of the model

5. Analysis of longitudinal data through non-linear models
   a. Introduction
   b. Graphical inspection of the data
   c. Estimation of a non-linear model
   d. Diagnosis of the model
   e. Solutions when model assumptions are not met
   f. Model selection
   g. Nonlinear Mixed model

Methodology

Theoretical sessions:

In these sessions the main concepts of each topic will be presented, as v

this task. The slides (created with R Markdown - which guarantees reproducibility of the results) will include the t

results and the conclusions drawn from them.

Practical sessions:

In these sessions some guided exercises will be proposed that the stude

database for the same real problem will be generated independently for each student. With this methodology, the

real data, regardless of whether he asks a classmate how to carry it out, since each student will have to analyze

Assistance to seminars:

Exceptionally, if there are certain time matches and if the students see it

Applied Statistics Service or by another research center close to the College. This assistance will not be compuls

students could see how the methodology they are learning is used in real studies and they could see how the wo

said investigations.

Individualized work

The solutions of all the practices that are carried out in person must be d

student will have a personalized database on the same problem to be addressed. In addition to these face-to-fac

at home and deliver the numerical solution, as well as the R code used to obtain said results. Both the classroom

continuous evaluation. The student will have access to all the teaching material from a teaching web in which the

ideally be resolved by their peers and that the teacher will supervise and / or solve, if necessary.

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 propo

Delivery of practices:

During the course the student will have to solve several practices in whic

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 acquir

questions about the models treated in class and R outputs similar to those obtained in the analyzes that have be

interpret the results.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Individual tasks	70%	0	0	8, 1, 2, 4, 5, 6, 12, 10, 11, 14, 13, 7
Test	30%	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.