Course guide
200609 - ATV - Lifetime Data Analysis

Unit in charge:	School of Mathematics and Statistics
Teaching unit:	715 - EIO - Department of Statistics and Operations Research.

Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Optional subject).

Academic year: 2023   **ECTS Credits:** 5.0   **Languages:** English

LECTURER

Coordinating lecturer: GUADALUPE GÓMEZ MELIS

Others: Primer quadrimestre:
GUADALUPE GÓMEZ MELIS - A
KLAUS GERHARD LANGOHR - A

PRIOR SKILLS

In order to follow the course successfully the student has to be familiar with the following concepts: estimation theory and confidence intervals, likelihood function, maximum likelihood estimation, regression models, hypothesis tests.
The student will have to use the R software for homework and data analysis.
Chapters 1 through 3 of the book "Principles of Statistical Inference" Cox, Cambridge University Press (2006) should be mastered.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
3. CE-2. Ability to master the proper terminology in a field that is necessary to apply statistical or operations research models and methods to solve real problems.
4. CE-3. Ability to formulate, analyze and validate models applicable to practical problems. Ability to select the method and / or statistical or operations research technique more appropriate to apply this model to the situation or problem.
5. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.
Translate to english
6. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.

Transversal:
2. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.

TEACHING METHODOLOGY

Lectures:
One hour and a half sessions in which the main concepts and topics are introduced. The lecturer will use a computer and the blackboard to introduce the course content. Emphasis is put on ideas and intuition. Topics are discussed from the point of view of real situations concerning clinical trials, epidemiological studies, sports data and other observational studies where the variable time until an event is of interest

Problem-solving sessions:
Incorporated into the practical sessions.

Laboratory sessions:
One hour and a half sessions held in the computer lab in which theoretical problems are tackled and exercises are carried out using computers.

LEARNING OBJECTIVES OF THE SUBJECT

Survival analysis is employed in many fields to analyze data representing the duration or elapsed time between two events. It is also known as event history analysis (in demography), lifetime data analysis, reliability analysis and time-to-event analysis. A key characteristic that distinguishes survival analysis from other areas of statistics is that survival data are usually censored, sometimes truncated and the normality hypothesis is inadequate. Censoring occurs when the information for some individuals is incomplete, what may happen for different reasons discussed in class.

The course Lifetime Data Analysis covers a series of procedures and techniques for analyzing censored and/or truncated data. While the course is focused on medical applications in public health, in epidemiology and in sports sciences, it also has direct applications to other disciplines such as economics, actuarial sciences, engineering and demography.

The aim of the course is to develop the core of survival analysis and to put into practice the knowledge acquired by means of the statistical software package R.

Abilities to be acquired:
* Identification of those situations or studies in which it is necessary to use Survival Analysis methodology. The ability to define the events and times relevant to each situation.
* Identification and knowledge of the different types of censoring and truncation. The ability to construct the likelihood in each case.
* Knowledge on the most common parametric models: Exponential, Weibull, Gamma, Gompertz, Lognormal and Log-Logistic. The ability to evaluate the most adequate model in a concrete example.
* The ability to obtain and interpret the Kaplan-Meier estimator, to know its most important properties and how to calculate estimators for the cumulative risk functions.
* Knowledge on how to present different hypothesis tests in order to compare two or more survival curves. The ability to select the most appropriate test according to the type of alternative hypothesis.
* Knowledge on how to use accelerated lifetime regression models: the Weibull and the log-logistic model. Knowledge of their relationships and differences.
* The ability to set out and interpret a proportional hazard model, as well as checking the goodness-of-fit by means of studying different residuals.

STUDY LOAD

Type	Hours	Percentage
Self study	80,0	64.00
Hours large group	30,0	24.00
Hours small group	15,0	12.00

Total learning time: 125 h

CONTENTS

Basic concepts and parametric models

Description:
Survival function. Hazard function.
Mean and median life
Principal parametric models.

Full-or-part-time: 12h 50m
Theory classes: 4h 30m
Self study : 8h 20m

Censoring and truncation

Description:
Different types of right censoring.
Left and interval censoring.
Building the likelihood function
Left truncation

Full-or-part-time: 11h 10m
Theory classes: 3h
Laboratory classes: 1h 30m
Self study : 6h 40m

One sample non-parametric inference

Description:
Kaplan-Meier estimator for the survival function.
Nelson-Aalen estimator for the cumulative risk function
Asymptotic Properties.
Confidence intervals and confidence bands.

Full-or-part-time: 30h 30m
Theory classes: 7h 30m
Laboratory classes: 3h
Self study : 20h

Two and K- sample comparison

Description:
Two and K-sample comparison
The (weighted) log-rank test.
Fleming-Harrington tests family.
Stratified tests

Full-or-part-time: 19h 10m
Theory classes: 6h
Laboratory classes: 1h 30m
Self study : 11h 40m

Parametric regression

Description:
Accelerated failure time models.
Log-linear, proportional hazards and proportional odds models.
Weibull regression model.
Log-logístic model.
General odds-rate regression model

Full-or-part-time: 20h 50m
Theory classes: 4h 30m
Laboratory classes: 3h
Self study : 13h 20m

Semi-parametric regression: Cox Model

Description:
Cox's regression model.
Partial likelihood function.
Inference within the Cox model.
Different types of residuals for the Cox model
Validation of the Cox model.

Full-or-part-time: 30h 30m
Theory classes: 7h 30m
Laboratory classes: 3h
Self study : 20h

GRADING SYSTEM

Assessment is based on the following:
* Problems solved and handed in throughout the course (3 sets) (25%)
* Case study with real data (25%)
* Final exam (50%)

EXAMINATION RULES.

The student will be informed at the beginning of the course on the dates of each deliverable.

BIBLIOGRAPHY

Basic:
- Lee, E.T.; Wang, J.W. Statistical methods for survival data analysis [on line]. 4th. Wiley, 2003 [Consultation: 05/07/2023]. Available on: https://onlinelibrary-wiley-com.recursos.biblioteca.upc.edu/doi/book/10.1002/0471458546. ISBN 9781118095027.
- Collett, D. Modelling survival data in medical research. 2nd ed. Chapman & Hall, 2003. ISBN 1584883251.
- Klein, John P.; Moeschberger, Melvin L. Survival analysis : techniques for censored and truncated data [on line]. 2nd ed. Springer, 2003 [Consultation: 05/07/2023]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/b97377. ISBN 978038795399.
- Smith, Peter J. Analysis of failure and survival data. Chapman and Hall, 2002. ISBN 1584880759.
- Kleinbaum, David; Klein, Mitchel. Survival analysis : a self-learning text. 3rd ed. Springer, 2012. ISBN 9781441966.

Complementary:
- Anderson, Stewart. Biostatistics : a computing approach. Boca Raton: CRC Press, cop. 2012. ISBN 9781584888345.
- Cox, D. R.; Oakes, D. Analysis of survival data. Chapman and Hall, 1984. ISBN 041224490X.

- Kalbfleisch, John D.; Prentice, R.L. The Statistical analysis of failure time data. 2nd ed. Wiley-Interscience, 2002. ISBN 047136357X.
- Lawless, Jerald F. Statistical models and methods for lifetime data. 2nd ed. 2003. ISBN 978-0471372158.
- Klein, John P. Handbook of survival analysis [on line]. Boca Raton: Taylor and Francis, cop. 2014 [Consultation: 05/07/2023]. Available on: https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=1563126. ISBN 9781466555662.
- O'Quigley, John. Proportional hazards regression [on line]. New York, NY: Springer New York, 2008 [Consultation: 05/07/2023]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-68639-4. ISBN 9780387686394.