Course guide
200641 - MLLG - Linear and Generalized Linear Models

Last modified: 11/04/2024

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: 2024 **ECTS Credits:** 5.0 **Languages:** English

LECTURER

Coordinating lecturer: VÍCTOR PEÑA PIZARRO

Others: Primer quadrimestre:
VÍCTOR PEÑA PIZARRO - A

PRIOR SKILLS

With respect to the Theory of Probability, the students should know the basic probability distributions, their main properties and the situations that they are able to model in an appropiate way. They also have to be familiarized with the main concepts of Statistical Inference corresponding to a first course of Statistics.

REQUIREMENTS

The student needs to know the base of statistics and probability and also an introduction on simple linear regresion. It is also necessary that the student has passed a course in matrix calculus. If the student also has some knowledge around ANOVA it would be good, but it is not mondatory.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
MESIO-CE4. CE-4. Ability to use different inference procedures to answer questions, identifying the properties of different estimation methods and their advantages and disadvantages, tailored to a specific situation and a specific context.
MESIO-CE3. 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.
MESIO-CE6. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.
MESIO-CE1. CE-1. Ability to design and manage the collection of information and coding, handling, storing and processing it.
MESIO-CE7. CE-7. Ability to understand statistical and operations research papers of an advanced level. Know the research procedures for both the production of new knowledge and its transmission.
MESIO-CE9. CE-9. Ability to implement statistical and operations research algorithms.
MESIO-CE8. CE-8. Ability to discuss the validity, scope and relevance of these solutions and be able to present and defend their conclusions.

Transversal:
CT3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.

CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT2. SUSTAINABILITY AND SOCIAL COMMITMENT: Being aware of and understanding the complexity of the economic and social phenomena typical of a welfare society, and being able to relate social welfare to globalisation and sustainability and to use technique, technology, economics and sustainability in a balanced and compatible manner.

TEACHING METHODOLOGY

The course will be taught in English. The course will be held in the first semester (S1) by means of two sessions per week. Usually, one session will be devoted to Theoretical questions and the other one to Practical. The practical sessions consist in the analysis of several data sets by means of the models presented in the theoretical sessions and statistical software R. In particular, we are going to use RStudio.

LEARNING OBJECTIVES OF THE SUBJECT

The main objective of this course is that the student acquires a good knowledge of Linear Models and Generalized Linear Models at both theoretical and practical levels. This knowledge will allow the student to analyze data sets and obtain conclusions.

Throughout the course, various data sets from very different fields will be analyzed, with the aim of being able to highlight some characteristics of a particular field. The knowledge imparted in this course will help the student to assimilate with greater ease and depth other subjects such as Longitudinal Models and Bayesian Analysis.

The knowledge and practice acquired in this course and subsequent modeling courses will allow the student to be able to collaborate with different research groups and advise them statistically.

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

Linear Model

Description:
Presentation and Linear Model.
1.1. Generalities. Objectives. Definition. Parameter Estimation and distribution. Residuals. Goodness of fit techniques. Checking the model hypothesis.
1.2. Analysis of Variance. One factor Anova: Parameter Estimation . Confidence Intervals for the means and means differences. Multiple comparisons. Two way ANOVA.
1.3. Multiple linear regressions: parameter estimation, determination coefficient, mean square error, confidence intervals for the parameters and estimations, model adequacy checking. Multiple regression: collinearity, causality, robust models and outliers detection. Parsimony principle. Anova Table. Common mistakes in regression.
1.4. Transformations.

Full-or-part-time: 12h
Theory classes: 7h
Laboratory classes: 5h

Exponential families

Description:
Definition. Canonical parameter. Parameter space. Examples and counter-examples. Different parametrizations of the same model. Maximum likelihood estimation.

Full-or-part-time: 5h
Theory classes: 3h
Practical classes: 2h

Generalized Linear models

Description:
3.1. Basic Concepts. Objectives. Definition. Link function and canonical link function. Parameter estimation and their asymptotic distribution. Goodness of fit measures: deviance, X^2 generalized Pearson statistic. Information criteria. Residuals.
3.2. Logit models: Estimation and testing. Parameter interpretation. Model selection Criteria for logistic regresion. Two dimensional tables and logistic regresion. Asymptotic results.
3.3. Log-linear models. Parameter estimation and test. Parameter interpretation. Models in two, three and higher dimensions. Residuals and outliers. Asymptotic results.
3.4 Negative binomial regression: Definition, overdispersion, interpretation, goodness of fit.
3.5 Multinomial regression: ordinal and nominal outcomes. Definition, interpretation, and goodness of fit.

Full-or-part-time: 28h
Theory classes: 20h
Practical classes: 8h

GRADING SYSTEM

T60% of the final grade will come from the final exam. This exam will contain a theoretical part and a data analysis part, both with the same weight. The remaining 40% of the final grade will come from the following activities:

1) Midterm exam with 10 short questions (20%).
2) A course project where the students model a dataset with RStudio (20%).

EXAMINATION RULES.

The exams will be closed book. For the midterm, the students can bring a calculator. For the data analysis portion of the final, the students will be able to use the course materials on Atenea, but nothing else.

BIBLIOGRAPHY

Basic:
- Seber, G.A.F. ; Lee, A. J. Linear regression analysis. Wiley, 2003. ISBN 0471415405.
- Dobson, J.A. An Introduction to generalized linear models. Chapman and Hall, 1990. ISBN 0412311100.
- Fox, J. Applied regression analysis and generalized linear models. Sage, 2008. ISBN 9780761930426.
- Fox, J. ; Weisberg, S. An R companion to applied regression. Sage, 2002. ISBN 9781412975148.

Complementary:
- McCullagh, P. ; Nelder, J.A. Generalized linear models. Chapman and Hall, 1989. ISBN 0412317605.
- Collet, D. Modelling binary data. Chaman and Hall, 2003. ISBN 1584883243.
- Lindsey, J. K. Applying generalized linear models. Springer, 1997. ISBN 0387982183.
- Montgomery, D. Design and analysis of experiments. 8 ed. Wiley, 2013. ISBN 1118097939.