200641 - MLLG - Linear and Generalized Linear Models

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
749 - MAT - Department of Mathematics

Academic year: 2017
Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Teaching unit Optional)
ECTS credits: 5  Teaching languages: Spanish, English

Teaching staff
Coordinator: MARTA PÉREZ CASANY
Others: Primer quadrimestre:
MARTA PÉREZ CASANY - A
JORDI VALERO BAYA - 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 appropriate way. They also have to be familiarized with the main concepts of Statistical Inference corresponding to a first course of Statistics.

Requirements
We start modelization from scratch, so there are no pre-requisites. Nevertheless, some knowledge about linear regression and/or ANOVA will help better understand the subject.

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-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
The main objectives of this subject are that the students acquire:

1) Deep knowledge of LINEAR MODELS. In particular of simple and multiple regression, ANOVA and ANCOVA.
2) Some skills on non-linear models that can be linearized.
3) Deep knowledge of GENERALIZED LINEAR MODELS. In particular of logistic regression, log-linear models, models for polytomous data, models for Gamma response.
4) Knowledge of modelling using QUASI-LIKELIHOOD.
5) Important level of practice dealing with real data.

This knowledge will be very useful when posteriorly, the students collaborate with research groups in different areas, with the objective of advise them in the statistical part.

These skills will allow the student:

1) To be able posteriorly to assimilate more easily other subjects as: LONGITUDINAL MODELS or BAYESIAN ANALYSIS
2) To be able to collaborate, at the end of the Master, with research groups of different kinds and give advice from the statistical point of view.
6) Ability in obtaining conclusions and explaining them.

The course will be held in the second semester (S2) in an intensive way, since it will last 7 weeks. Each week there will be two sessions of three hours divided in two parts, with a break of 15 minutes. The first part corresponds to the theory session and will take place in a normal room. The second part will take place in a computer room since it consists in the analysis of some data sets by means of the statistical software R.

## Learning objectives of the subject

The main objectives of this subject are that the students acquire:

1) Deep knowledge of LINEAR MODELS. In particular of simple and multiple regression, ANOVA and ANCOVA.
2) Some skills on non-linear models that can be linearized.
3) Deep knowledge of GENERALIZED LINEAR MODELS. In particular of logistic regression, log-linear models, models for polytomous data, models for Gamma response.
4) Knowledge of modelling using QUASI-LIKELIHOOD.
5) Important level of practice dealing with real data.

This knowledge will be very useful when posteriorly, the students collaborate with research groups in different areas, with the objective of advise them in the statistical part.

These skills will allow the student:

1) To be able posteriorly to assimilate more easily other subjects as: LONGITUDINAL MODELS or BAYESIAN ANALYSIS
2) To be able to collaborate, at the end of the Master, with research groups of different kinds and give advice from the statistical point of view.
6) Ability in obtaining conclusions and explaining them.

## Study load

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group: 30h</th>
<th>24.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours small group: 15h</td>
<td>12.00%</td>
<td></td>
</tr>
<tr>
<td>Self study: 80h</td>
<td>64.00%</td>
<td></td>
</tr>
</tbody>
</table>

The course will be held in the second semester (S2) in an intensive way, since it will last 7 weeks. Each week there will be two sessions of three hours divided in two parts, with a break of 15 minutes. The first part corresponds to the theory session and will take place in a normal room. The second part will take place in a computer room since it consists in the analysis of some data sets by means of the statistical software R.
# Content

<table>
<thead>
<tr>
<th>Linear Model</th>
<th>Learning time: 18h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 10h 30m</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 7h 30m</td>
</tr>
</tbody>
</table>

**Description:**

Presentation and Linear Model.
1.4. Transformations to obtain linearity, normality and/or homocedasticity. Non linear models than can be linearized.

<table>
<thead>
<tr>
<th>Exponential families</th>
<th>Learning time: 6h 45m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 3h 45m</td>
</tr>
<tr>
<td></td>
<td>Practical classes: 3h</td>
</tr>
</tbody>
</table>

**Description:**

Qualification system

The 60% of the Final mark will come from the Final Exam. This exam will contain a theoretical as well as a practical part, both with the same weight. The remaining 40% will come from the activities realized during the course. The activities jointly with their weights are the following:

1) Reading, report and oral presentation of a scientific paper (10%).
2) Mini Exam composed by 10 short questions (10%).
3) Two deliveries in which the student will need to model a set of data with R (20%).

Regulations for carrying out activities

The Mini Exam and the Final Exam will be closed book, but the students might need to bring calculator and statistical tables.
Bibliography

Basic:

Fox, J.; Weisberg, S. An R companion to applied regression. sage, 2011.

Complementary: