Universitat Autònoma de Barcelona

2021/2022

Mathematical Statistics

Code: 106081
ECTS Credits: 6

Degree	Type	Year	Semester
2500149 Mathematics	OT	4	0

The proposed teaching and assessment methodology that appear in the guide may be subject to changes as a result of the restrictions to face-to-face class attendance imposed by the health authorities.

Contact

Name: Mercè Farré Cervelló
Email: Merce.Farre@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Prerequisites

The competencies in algebra, analysis and probability and statistics of the first cycle of mathematics are assumed.

Objectives and Contextualisation

The students will learn to formalize, analyze and validate models that attempt to assess the relationships between different variables under uncertainty conditions whithin the mathematical statistics setting, in order to provide confidence intervals for the parameters'model and perform statistical hypotheses tests.

The goal of a regression model is to explain the mean behaviour of a response variable in terms of other variables related to it. Given a model, predictions and residuals can be obtained and analyzed, analysis that will be translated into decisions at an experimental level. The students must be conscious of the constraints in each mathematical model and select which model behaves better. Thus, they must know how to adjust, validate and compare various linear models and select the best set of regressors in each case.

Some extensions of the linear model are also introduced, such as generalized linear models, polynomial or nonlinear models, for example, as they extend the scope of modeling and allow constraints to be lowered. The general linear model is a theoretical framework that allows to formulate the techniques of analysis of the variance and the design of experiments within the linear model.

With this course, students will be able to explore and validate the theoretical properties of the general linear model, learn some extensions, and be trained to model data with free software. They will needto understand in depth the importance of the most important theorems in this area, as well as their proof.

Competences

- Actively demonstrate high concern for quality when defending or presenting the conclusions of ones work.
- Effectively use bibliographies and electronic resources to obtain information.
- Generate innovative and competitive proposals for research and professional activities.

- Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Students must develop the necessary learning skills to undertake further training with a high degree of autonomy.
- Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
- Understand and use mathematical language.

Learning Outcomes

1. Actively demonstrate high concern for quality when defending or presenting the conclusions of ones work.
2. Effectively use bibliographies and electronic resources to obtain information.
3. Generate innovative and competitive proposals for research and professional activities.
4. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
5. Students must develop the necessary learning skills to undertake further training with a high degree of autonomy.
6. Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
7. Understand abstract language and understand in-depth demonstrations of some advanced theorems of probability and statistics.

Content

Preliminaries

• The simple linear model: least squrares, maximum likelihood and other estimation methods.

• Multivariate Gaussian distributions and related laws.

The multiple linear model

• The linear model. Normal equations. Properties of the coefficients' and variance estimators. BLUE. Goodness of fit indicators.

- Estimation of the mean response and prediction of new observations.

• Sum of squares decompositions and distributions. Hypothesis tests and confidence regions. The Cochran theorem.

• Model diagnostics. Transformations.

• Outliers and influential observations.

- The multicolinearity problem. The bias problem. Model selection criteria.

Design of experiments, anova and the general linear model

• One-way analysis of variance. Multiple comparisons.

• Analysis of the variance with several factors. Interactions.

• The design of experiments setting.

• The response surface models.

- Dummy variables in regression and the general linear model.

Certain extensions of the linear model

- Random effects models. Repeated measures models.
- Generalized linear models: binomial, Poisson, etc.
- Nonlinear regression.

Methodology

The statistical models and their corresponding assumptions and properties are introduced in the theoretical sessions. Emphasis will be placed on rigor in the proofs as well as on the applicability and interpretation of the methods.

The discussion will be encouraged in the classroom and theoretical problems will be proposed to deepen the topics. Problems, and practical exercises to be performed with free software R will be proposed, with the aim that students will be able to model data. Some sections of the course will be developed by students in the form of work and will be a written as a short report and presented to the classroom.

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			
Computer work	24	0.96	1, 3, 4, 2
Problems sessions	6	0.24	7, 4, 2
Theoretical classes	30	1.2	7, 2
Type: Autonomous			
Personal work	80	3.2	3, 4, 2

Assessment

The evaluation scheme is as follows:

NC = 0.3 * P1 + 0.4 * P2 + 0.15 * Tb + 0.15 * Lli

P1: First partial exams (30%) = theory and problems (15%) + computer test (15%).

P2: Second partial exams (40%) = theory and problems (20%) + computer test (20%).

Tb: Personal project (15%).

Lli: Delivery of solved problems and practical exercices (15%).

Besides that, the students will have the option of taking an additional recovery exam (RE) with the same format, to recover only the P1+P2 amount.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
First partial exam	0,2	4	0.16	7, 4, 2
Oral exposition of a report	0,2	1	0.04	1, 3, 6, 5, 4, 2
Second partial exam	0,3	4	0.16	1, 4, 2
Tasks delivery	0,3	1	0.04	3, 4, 2

Bibliography

- Peña, D.; Regresión y Diseño de Experimentos. Alianza Editorial, 2002.
- Rao, C. B., Touttenburg, H., Shalabh, Heumann, C.; Linear Models and Generalizations. 3rd edition, Springer, 2008.
- Rawlings J. O, Pantula S. G , Dickey D. A.; Applied Regression Analysis. A Research Tool. Second Edition, Springer, 1999.
- Rencher, A.C., Schaalje, G.B.; Linear Models in Statistics. Wiley-Interscience, 2008.
- Seber, G., Lee, A.; Linear Regression Analysis. Wiley Series in Probability and Statistics, 2003.
- Hay-Jahans C.; An R Companion to Linear Statistical Models. Chapman and Hall, 2012.
- Faraway, J.; Linear Models with R. Chapman&Hall/CRC, 2005.
- Faraway, J.; Extending the linear model with R. Chapman&Hall/CRC, 2006.
- Vikneswaran; An R Companion to Experimental Design. https://cran.r-project.org/doc/contrib/Vikneswaran-ED_companion.pdf

Complementary references

- McCullagh, P., Nelder, J. A.; Generalized Linear Models. Chapman&Hall, 1989.
- Clarke, B. R.; Linear Models. The theory and applications of Analysis of Variance. Wiley Series in Probability and Statistics, 2008.
- Sen, A., Srivastava, M.; Regression Analysis. Theory, Methods and Applications. Springer, 1990.
- Carmona, F.; Modelos Lineales. Universitat de Barcelona, 2005.
- Christensen, R.; Advanced Linear Modelling. Springer, 2001.
- Christensen, R.: Log-Linear Models. Springer, 1990.
- Draper, N., Smith, H.; Applied regression Analysis. Wiley, 1998.
- Chatterjee, S. & Price, B.; Regression Analysis by Example. Wiley-Interscience, third edition, 2000.
- Scheffé, H.; The Analysis of Variance, 1999.
- Montgomery, D.C., Peck, E., Vining, G.; Introduction to Linear Regression. Wiley Series in Probability and Statistics, 2001.

Software

Free software R and Rstudio.