Universitat Autònoma de Barcelona

2020/2021

Linear Models 2

Code: 104861
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	OB	3	1

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: Natalia Isabel Vilor Tejedor
Email: NataliaIsabel.Vilor@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

Basic knowledge of descriptive and inferential statistics. A previous course of Linear Models is required.

Objectives and Contextualisation

This course is based on supervised learning, with a focus on regression and classification methods. The syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis; model selection and regularization methods (ridge and lasso); nonlinear models such as splines and generalized additive models.

Competences

- Analyse data using statistical methods and techniques, working with data of different types.
- 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.
- Design a statistical or operational research study to solve a real problem.
- Formulate statistical hypotheses and develop strategies to confirm or refute them.
- 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.
- Select and apply the most suitable procedures for statistical modelling and analysis of complex data.
- Select statistical models or techniques for application in studies and real-world problems, and know the tools for validating them.
- 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.
- Summarise and discover behaviour patterns in data exploration.
- Use quality criteria to critically assess the work done.

Learning Outcomes

1. Analyse data through inference techniques using statistical software.
2. Analyse data using the generalised linear model.
3. Analyse data using the model of linear regression.
4. Analyse the residuals of a statistical model.
5. Choose the relevant explanatory variables.
6. Compare the degree of fit between several statistical models.
7. Critically assess the work done on the basis of quality criteria.
8. Detect and contemplate interactions between explanatory variables.
9. Detect and respond to colinearity between explanatory variables.
10. Draw conclusions about the applicability of models with the use and correct interpretation of indicators and graphs.
11. Establish the experimental hypotheses of modelling.
12. Identify response distributions with the analysis of residuals.
13. Identify sources of bias in information gathering.
14. Identify the response, explanatory and control variables.
15. Identify the stages in problems of modelling.
16. Identify the statistical assumptions associated with each advanced procedure.
17. Make effective use of references and electronic resources to obtain information.
18. Make slight modifications to existing software if required by the statistical model proposed.
19. Measure the degree of fit of a statistical model.
20. Predict responses, compare groups (causal value) and identify significant factors.
21. Reappraise one's own ideas and those of others through rigorous, critical reflection.
22. 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.
23. 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.
24. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
25. Summarise and interpret the results from classic and generalised linear models and from non-linear models on the basis of the objectives of the study.
26. Use a range of statistical software to adjust and validate linear models and their generalisations.
27. Use graphics to display the fit and applicability of the model.
28. Use logistic regression to solve classification problems.
29. Validate the models used through suitable inference techniques.

Content

1. Linear Regression
Simple Linear Regression
Multiple Linear Regression
Extension of the Linear Models

2. Classification
Overview of Classification.
Logistic Regression: The Logistic Model. Estimating the Regression Coefficients. Predictions.
Multiple Logistic Regression
Linear Discriminant Analysis.
Quadratic Discriminant Analysis.

3. Linear Model Selection and Regularization
Subset Selection: Best Subset Selection, Stepwise Selection, Optimal model selection.
Shrinkage Methods: Ridge Regression and LASSO regression. Selecting the Tuning Parameter
Dimension Reduction Methods: Principal Component Analysis and Partial Least Squares

4. Moving Beyond Linearity
Polynomial Regression
Step-wise Regression
Splines
Generalized Additive Models

*Unless the requirements enforced by the health authorities demand a prioritization or reduction of these contents.

Methodology

The course material (theory notes, lists of problems and statements of practice) will be available at the virtual campus, progressively throughout the course.

*The proposed teaching methodology may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Computer Practices	50	2	3, 2, 1, 21, 7, 11, 27, 15, 16, 18, 23, 25, 17, 28
Theory	50	2	3, 2, 1, 4, 21, 7, 6, 8, 9, 11, 10, 27, 12, 13, 15, 16, 14, 19, 18, 20, 24, 22, 23, 5, 25, 17, 28, 26, 29
Type: Supervised			
problems / exercises to solve	16	0.64	1, 21, 7, 10, 23, 25, 17, 26
Type: Autonomous			
Preparation for the exam	10	0.4	1, 21, 7, 16, 24, 25

Assessment

PR: Practices. PR score: 4 points out of 10. Not recoverable.

P1: Test 1 (theory, problems and practices, online). P1 score: 2 points out of 10.

P2: Test 2 (theory, problems and practices). P2 score: 4 points out of 10.

It is necessary to obtain a minimum score of 3.5 points in each exam. The final grade will be: Final grade = PR + P1 + P2.

In January there will be a final test, PF, which allows the recovery of P1 and P2 (6 points out of 10). Then, the final grade will be: Final grade = PR + PF.

*Student's assessment may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Practices	40%	16	0.64	1, 21, 7, 11, 27, 15, 16, 18, 24, 22, 23, 25, 17, 26
Test 1	20%	4	0.16	3, 2, 1, 4, 21, 7, 6, 8, 9, 11, 10, 27, 12, 13, 15, 16, 14, 19, 18, 20, 24, 22, 23, 5, 25, 17, 28, 26, 29
Test 2	40%	4	0.16	3, 2, 1, 4, 21, 7, 6, 8, 9, 11, 10, 27, 12, 13, 15, 16, 14, 19, 18, 20, 24, 22, 23, 5, 25, 17, 28, 26, 29

Bibliography

Montgomery, D. Peck, A. Vining, G.; *Introduction to Linear Regression Analysis*. Wiley. 2001.

Christopher Hay-Jahans; *An R Companion to Linear Statistical Models*. Chapman and Hall, 2012.

John Fox and Sandord Weisberg; *An R Companion to Applied Regression*, 2nd edition, Sage Publications, 2011.

Daniel Peña; *Regresión y diseño de Experimentos*, Alianza Editorial (Manuales de Ciencias Sociales), 2002.

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani; An Introduction to Statistical Learning, Springer texts in Statistics, 2013.