UAB
Universitat Autònoma de Barcelona

2020/2021

Linear Models 1

Code: 104860
ECTS Credits: 6

Degree	Type	Year	Semester
2503852 Applied Statistics	OB	2	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: 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

Fundamentals of descriptive and inferential statistics and probabilities, as well as knowing the rudiments of programming with the R language.

Objectives and Contextualisation

The objective of the course is to study the modeling and analysis of data using the theory of Linear Models, as well as applications in various fields (economics, health, engineering, and science in general). The methods and techniques are introduced based on examples and developed by means of the resolution of a number of proposed problems together with computer work with to be developed in the R environment. First, the simple regression model is presented due to its numerous applications and because it is a good prologue to the understanding of the multiple model. The multiple regression includes some variants (polynomial, with interactions, using fictitious regressive variables, etc.) and constitutes the second part of the course. In all the modeling procedures, the goodness of fit and the correct model' specification, the theoretical assumptions and the detection of "special" (anomalous and influential) data are analyzed, and possible solutions are proposed in the case that a flagrant violation of the model hypotheses is found.

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 model of block analysis of variance.
3. Analyse data using the model of covariance analysis.
4. Analyse data using the model of variance analysis for one or more factors.
5. Analyse data using the model of variance analysis with nested factors.
6. Choose the relevant explanatory variables.
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. Establish the experimental hypotheses of modelling.
11. Identify sources of bias in information gathering.
12. Identify the stages in problems of modelling.
13. Identify the statistical assumptions associated with each advanced procedure.
14. Make effective use of references and electronic resources to obtain information.
15. Make slight modifications to existing software if required by the statistical model proposed.
16. Plan studies based on time series.
17. Predict responses, compare groups (causal value) and identify significant factors.
18. Reappraise one's own ideas and those of others through rigorous, critical reflection.
19. 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.
20. 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.
21. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
22. Use a range of statistical software to adjust and validate linear models and their generalisations.
23. Use specific packages for experimental design.
24. Use statistical software to calculate sample size.
25. Validate the models used through suitable inference techniques.

Content

1. The simple linear regression model.

- Introduction to regression: Exploring data.

- Simple linear regression: Model, hypotheses and parameters.

- Punctual estimation: Least squares and the maximum likelihood methods.

- Inference about the parameters under the Gauss-Markov hypothesis: Intervals and tests.

- New observations: The confidence interval for the mean response and the prediction intervals. Simultaneous inferences. Confidence and prediction bands.

- Analysis of the variance (ANOVA) in simple regression.

- Model diagnostics: Graphical evaluation of the linearity and the model hypotheses through the analysis of the residuals. The lack of fit test.

- Anomalous and influential data.

2. Multiple linear regression

- Previous steps in multiple regression: Exploration of data with multidimensional visualizing tools.

- Model and estimators of the coefficients by least squares. Interpretation of the coefficients in the multiple linear model.

- Laws of estimators of coefficients, predictions and residuals: application of the properties of idempotent matrices.

- Inference in the multiple linear model. The model anova.

- Linear constraints on the coefficients: The incremental variability principle.

- Discussion on the model hypotheses: Analysis of the residuals. Box-Cox transformations.

- The multicollinearity problem: Detection and solutions.

- Dummy variables in regression.

- Variables selection: Mallows Cp statistic, cross validation and automatic stepwise selection procedures.

Methodology

The subject has two weekly hours of theory and problems, where linear methods and tools are introduced and analyzed. Problem lists will be supplied along the course, to be delivered. Practical sessions will be carried out using the R programming language. Tasks to be delivered are proposed related to the theoretical exercices and to the computer practical work. The student will also perform extra autonomous consisting of bibliographical research and exams preparation.

The course material (theory notes, lists of problems and computer tasks) will be available in the *moodle* classroom.

The gender perspective goes beyond the contents of courses, since it implies also a revision of teachingmethodologies and interactions between students and lecturers, both inside and outside the classroom. In this sense, participative teaching methodologies that give rise to an equality environment, less hierarchical in theclassroom, avoiding examples stereotyped in gender and sexist vocabulary, are usually more favorable to the full integration and participation of female students in the classroom. Because of this, their effective implementation will be attempted in this course.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Supervised computer sessions	26	1.04	1, 13, 15, 6, 14, 22
Theoretical classes	26	1.04	3, 2, 5, 4, 18, 7, 10, 11, 12, 13, 6, 25
Type: Autonomous			
Computer work	32	1.28	3, 2, 5, 4, 1, 8, 9, 24, 11, 16, 17, 19, 20, 6, 23, 22, 25

Personal work	36	1.44	18, 7, 14
Problem solving	18	0.72	4, 10, 11, 12, 13, 21, 19, 6, 25

Assessment

PR: Delivery of the theoretical and practical (with R) exercises. Maximum PR rating: 2 points. This part is not recoverable.

P1: Partial test of simple regression (theory, exercises, and practices). Maximum rating of P1: 3 points.

P2: Multiple regression partial test (theory, exercises and practices). Maximum rating of P2: 5 points.

The course grade will be calculated: NC = PR + P1 + P2. It is mandatory for NC be equal to or greater than 5 and that the grades of each partial be greater than or equal to 3.5 (out of 10).

At the end of the semester there will be a recovery test that will be a synthesis test, PS, (theory, exercises and practices) of the contents of the entire course with a maximum score of 8 points, by students who have not passed by course or want to improve the note. Only students who have participated in 2/3 of the evaluation activities may be submitted to the synthesis test.

The final grade of those presented to the synthesis test will be calculated: NF = PR + max (PS, P1 + P2).

Honor grades will be granted at the first complete evaluation. Once given, they will no be withdrawn even if another student obtains a larger grade after consideration of the PS exam.

Attention: "Without prejudice to other disciplinary measures deemed appropriate, and in accordance with current academic regulations, will be scored with a zero the irregularities committed by the student that may lead to a variation of the rating of an evaluation act. Therefore, plagiarizing, copying or letting a practice copy or any other evaluation activity involve suspending with a zero and cannot be recovered in the same academic year. If this activity has a minimum associated score, then the subject will be suspended. "

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Final test	80% (recovery partial exams)	4	0.16	4, 1, 10, 11, 12, 13, 15, 6, 22, 25
Partial exam 1	30%	4	0.16	2, 4, 1, 10, 24, 13, 17, 22, 25
Partial exam 2	50%	4	0.16	3, 2, 5, 4, 1, 8, 9, 10, 11, 13, 15, 16, 17, 6, 22, 25
Tasks delivery	20%	0	0	2, 5, 4, 1, 18, 7, 8, 9, 24, 15, 17, 21, 19, 20, 6, 14, 23, 22, 25

Bibliography

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

Clarke, B.R.; Linear Models:The Theory and Applications of Analysis of variance. Wiley, 2008.

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

Fox, J. and Weisberg, S.; An R Companion to Applied Regression.Sage Publications, 2nd edition, 2011.

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

Complementary references:

Sen, A., Srivastava, M.;Regression Analysis: Theory, Methods and Applications. Springer, 1990.

Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman; .Applied Linear Models. Irwin (4th edition), 1996.

Faraway, J.; Linear Models with R. Chapman&Hall/CRC (2nd ed), 2014.

Rao, C. R., Toutenburg, H., Shalabh, Heumann, C; Linear Models and generalizations. Springer, 2008.