UAB
Universitat Autònoma de Barcelona

Linear Models 1

Code: 104860
ECTS Credits: 6

2024/2025

Degree	Type	Year
2503852 Applied Statistics	OB	2

Contact

Name: Maria Merce Farre Cervello

Email: merce.farre@uab.cat

Teaching groups languages

You can view this information at the end of this document.

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 aim of the subject is the study of Linear Models, as well as the applications in various fields. The methods and techniques are introduced from examples and worked through the solution of the proposed problems and computer practices designed to be executed with the R language. First, the simple regression model is presented because it has numerous applications and because it is a good introduction to understanding the multiple model. The multiple regression model, expressed matrixically and which includes some variants (polynomial, dummy regressors, interactions, etc.), constitutes the second part of the course. Using residual analysis as the primary tool, model fit and correct specification, hypothesis fulfillment, and detection of "special" (anomalous and/or influential) data are analyzed. Finally, topics of particular relevance are addressed, such as multicollinearity and variable selection.

Learning Outcomes

1. CM09 (Competence) Assess the suitability of the models with the correct use and interpretation of indicators and graphs.
2. CM09 (Competence) Assess the suitability of the models with the correct use and interpretation of indicators and graphs.
3. CM10 (Competence) Modify the existing software if required by the statistic model, or create new software, if necessary.
4. KM12 (Knowledge) Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.
5. KM12 (Knowledge) Provide the experimental hypotheses of modelling, considering the technical and ethical implications involved.
6. KM13 (Knowledge) Detect interactions, co-linearity and importance between explanatory variables.
7. SM11 (Skill) Analyse the residuals of a statistical model.
8. SM12 (Skill) Interpret the results obtained to formulate conclusions about the experimental hypotheses.
9. SM13 (Skill) Compare the degree of adjustment between diverse statistical models.

10. SM14 (Skill) Use graphs to visualise the fit and suitability of the model.

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.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Supervised computer sessions	26	1.04	

		26	1.04
Theoretical classes		26	1.04
Type: Autonomous			
Computer work		32	1.28
Personal work		36	1.44
Problem solving		18	0.72

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.

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.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Final test	80% (recovery partial exams)	4	0.16	CM09, KM13, SM11, SM12, SM13, SM14
Partial exam 1	30%	4	0.16	CM09, KM13, SM11, SM12, SM13, SM14
Partial exam 2	50%	4	0.16	
Tasks delivery	20%	0	0	CM09, CM10, KM12, KM13, SM11, SM12, SM13, SM14

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.

Unique assessment

The single assessment will be a synthesis test of the skills of the two partials, based on: (1) An exam with theory and practical questions (weight: 50%). (2) A practice test in front of the computer (weight: 40%). (3) The delivery of the scheduled tasks that are indicated, with the possibility of the professor asking the student to explain details of these deliveries (weight: 10%).

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. "

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.

N. R. Mohan Madhyastha; S. Ravi; A. S. Praveena. A First Course in Linear Models and Design of Experiments. 2020.
https://link-springer-com.are.uab.cat/content/pdf/10.1007%2F978-981-15-8659-0.pdf

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.

Software

Free software: R and RStudio.

Language list

Name	Group	Language	Semester	Turn
(PLAB) Practical laboratories	1	Catalan	second semester	afternoon
(PLAB) Practical laboratories	2	Catalan	second semester	afternoon
(TE) Theory	1	Catalan	second semester	afternoon