VNIVERSITAT DE VALÈNCIA

Course Guide
36417 Linear models

COURSE DATA

Data Subject	
Code	36417
Name	Linear models
Cycle	Grade
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	2	Second term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	6 - Statistical Modelling	Obligatory

Coordination

Name	Department
AMOROS SALVADOR, RUBEN	130 - Statistics and Operational Research

SUMMARY

Linear Models is a compulsory subject taught in the second semester of the second year of the Degree in Data Science. Following on from the first year courses on Probability and Statistical Inference taught in the first semester of the second year, the aim of this course is to provide students with basic theoretical knowledge of statistical modelling as well as practical knowledge of its use in common experimental situations.

The theory classes will be taught in Spanish. The language for the practical and laboratory classes will be stated in the course guidelines available on the website for this degree.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

Knowledge of Linear Algebra, Probability and Statistical Inference is required.

OUTCOMES

1406 - Degree in Data Science

- (CG01) Knowledge of basic subjects and technologies that enable students to learn new methods and technologies, and to provide them with versatility to adapt to new situations.
- (CT01) To be able to access (bibliographical) information tools and appropriately use them in the development of their daily tasks.
- (CE09) To methodologically know and apply the concepts and techniques of probability and statistics necessary for the extraction of useful knowledge from data analysis.
- (CE15) Ability to model and analyse the uncertainty in data-based studies, as well as to know how to interpret and contextualise the results obtained.
- (CB4) Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.
- (CB5) Students must have developed the learning skills needed to undertake further study with a high degree of autonomy.

LEARNING OUTCOMES

Understand and apply linear regression models with one or more input variables. (Competences CB4, CB5, CG01, CT01, CE09, CE15)

Understand and apply linear classification models. (Competences CB4, CB5, CG01, CT01, CE09, CE15)

Understand and apply classical procedures for selecting variables by stages, regularization methods and dimensionality reduction. (Competences CB4, CB5, CG01, CT01, CE09, CE15)

DESCRIPTION OF CONTENTS

1. Simple linear regression

1.1. Simple linear regression: least squares and the lineal model
1.2. Parameter estimation and prediction
1.3. Model diagnosis: normality, homoscedasticity, and independency
1.4. Detection of influential observations and outliers.

2. Multiple linear regression

2.1. Multiple linear regression: Matrix representation
2.2 Models with fixed and random effects. Interaction
2.3. Identifiability conditions and coefficient estimation
2.4. Hypothesis testing and nested model comparison

3. Model selection

3.1. Fitting measures
3.2. Criteria for selecting models
3.3. Validation and cross validation
3.4. Bootstrap

4. Supervised classification and GLMs

4.1 Supervised classification
4.2 Generalised linear models: logistic and Poisson regression
4.3 Linear discriminant analysis

5. Regularization methods and dimensionality reduction

5.1. Ridge regression
5.2. Lasso
5.3. Partial least squares
5.4. Principal components regression

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	32,00	100
Laboratory practices	20,00	100
Classroom practices	8,00	100
Development of group work	30,00	0
Study and independent work	20,00	0
Preparation of evaluation activities	10,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	10,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

MD1 - Theoretical classes. Expository development of the subject (CG01). Students will be encouraged to participate in solving specific problems.

The topics discussed on the course are designed to encourage student participation.

MD2 - Practical classes. Students' work will focus on solving problems, completing exercises, and analyzing case studies (CB04, CB05, CT01, CE09, CE15).

The practical activities will complement the theoretical explanations, show students how to apply the basic concepts in real data analysis situations, and emphasize the advantages and limitations of each modelling technique.

MD4 - Work in the computer classroom. Students will learn by carrying out activities in the computer classroom either individually or in small groups (CB04, CB05, CT01, CE09, CE15).

As well as completing classroom activities, students will work on practice-related activities, prepare for classes and study for exams (CG01). Some of these tasks will be completed individually but others will require students to work in small groups in order to improve their ability to work as members of a team.

The University of Valencia's virtual classroom (*Aula Virtual*) will be used to communicate with students. Students will also be able to access the learning materials used in class and upload their assignments via the virtual classroom.

EVALUATION

At the first examination sitting, the course will be evaluated by the following three components:

SE1: An exam at the end of the semester will consist of theoretical and practical questions and problems (assessment of competencies CB04, CB05, CG01, CE09, CE15). This component will account for 50% of the final grade for this course. A minimum score of 4.5 out of 10 must be obtained in this section for the evaluation of the other sections to be taken into account in the final assessment.

SE2 – The computer practice assignments will be evaluated based on the student's reports and/or oral presentations (assessment of competencies CB04, CT01, CG01, CE09, CE15). This component of the evaluation will account for 40% of the final grade for this course. A minimum score of 4.5 out of 10 must be obtained in this section for the evaluation of the other sections to be taken into account in the final assessment.

SE3 – Continuous assessment will be based on the student's participation and degree of involvement in the learning process, attendance at face-to-face activities, and solution of questions and problems set periodically. This component of the evaluation will account for 10% of the final grade for this course and cannot be retaken (assessment of competencies CB05, CG01, CE09, CE15).

At the second examination sitting, the exam outlined in section SE1 above will be repeated. The computer practice tasks outlined in section SE2 above will be re-evaluated via an individual practical examination carried out in identical conditions to those of a practical assignment but with a time limit and limited access to support materials. For the continuous assessment component outlined in section SE3 above, the grade obtained by the student during the course will be used since this component cannot be retaken. All three components will be weighted by the same percentages as at the first examination sitting and the same minimum grade of 4.5 out of 10 will still apply for sections SE1 and SE2.

In all cases the evaluation system will be governed by the University of Valencia's regulations on grading and assessment for bachelor's degrees and master's degrees, which is available at:

http://www.uv.es/graus/normatives/2017_108_Reglament_avaluacio_qualificacio.pdf

REFERENCES

Basic

- P.K. Dunn and G.K. Smyth (2018) Generalized Linear Models with Examples in R. Springer
- J.J. Faraway (2014) Linear Models with R (2nd edition). Taylor and Francis

- G. James, D. Witten, T. Hastie and R. Tibshirani (2013) An Introduction to Statistical Learning with Applications in R. Springer

Additional

- J.J. Faraway (2016) Extending the Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models (2nd edition). Taylor and Francis
- A. Agresti (2015) Foundations of Linear and Generalized Linear Models. Wiley
- M. H. J. Gruber and S.R. Searle (2016). Linear Models. Wiley