COURSE DATA

DATA SUBJECT

Code: 34177
Name: Statistical modelling
Cycle: Undergraduate Studies
ECTS Credits: 6
Academic year: 2025-26

STUDY (S)

Degree	Center	Acad. year	Period
1107 - Degree in Mathematics	Facultat de Ciències Matemàtiques	4	Second quarter

SUBJECT-MATTER

Degree	Subject-matter	Character
1107 - Degree in Mathematics	Models of statistics and operations research	ELECTIVES

COORDINATION

ARMERO CERVERA MARIA CARMEN

SUMMARY

This course is intended as a basic introduction to statistical modeling. It is the natural continuation of the mathematical statistics course. Once students know the language of statistics, probability, and have learned the basic concepts and procedures of statistical inference, such as estimation and hypothesis testing, it is time to learn linear models. These are the most paradigmatic statistical models that allow connecting the theory of mathematical statistics with the reality of data. This is a traveling subject. It will move between the theoretical world of linear models, with linear algebra as an essential tool and facilitator of theoretical developments and the world of data and real problems. Both worlds are connected. The main objective of this course is to learn from both worlds, to value the tools provided by the world of linear models theory to go down to the world of data and to be able to provide serious and rigorous answers to the scientific questions generated by real problems with data and uncertainty.

PREVIOUS KNOWLEDGE

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

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

OTHER REQUIREMENTS

Have studied the subjects of Probability (2nd year) and Statistics Mathematics (3rd year).

COMPETENCES / LEARNING OUTCOMES

1107 - Degree in Mathematics

Adapting to new situations.

Apply the knowledge in the professional world.

Argue logically in decision-making.

Capacity of abstraction and modeling.

Expressing mathematically in a rigorous and clear manner.

Knowing the time and the historical context in which occurred the great contributions of women and men in the development of mathematics.

Learn autonomously.

Participate in the implementation of software and learn mathematical software.

Reason logically and identify errors in the procedures.

Visualize and interpret the solutions obtained.

DESCRIPTION OF CONTENTS

1. Introduction to regression models

Reminder of basic elements of statistical inference. Simple linear regression model without intercept. Least-squares estimation. Sampling distribution of model estimators. Point and interval estimation. Hypothesis testing.

2. The simple linear regression model.

The simple linear regression model. Estimation of model parameters. Bivariate normal distribution. Properties of the estimators. Inference about the model parameters. The regression contrast and the ANOVA table. Prediction. Correlation coefficient and coefficient of determination.
Important: This is a theoretical-practical topic that will be developed jointly by the students and the teacher in class simultaneously with Session 3. It will allow us to reinforce the concepts of Session 1 and will later serve as a theoretical example of the material of Session 3.

3. The general linear model.

Definition. Multivariate normal distribution. Design matrix. Identifiability. Regression models and analysis of variance models. Least squares estimation. Distribution of model coefficient estimators. Hypothesis testing. Influence matrix, residuals and fitted values. Gauss-Markov theorem. Geometry of linear models.

4. Linear models in Dataland.

Linear models and R. Model fitting and revision: residuals, outliers, influential observations, heteroscedasticity, etc. Confidence and prediction intervals. Col-linearity, confounding and causality. Identifiability. Model selection. Factors and interaction.

5. Generalized linear models.

Exponential family of distributions. Generalized lineal model (GLM). Model fitting. Asymptotic distribution of the estimators of the regression coefficients. Model comparison. Canonical link functions. Residuals. Quasi-likelihood. The geometry of linear modelling. Logistic regression and Poisson regression. GLMs and R.

WORKLOAD

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	37,50
Other activities	7,50
Computer classroom practice	15,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	5,00
Independent study and work	30,00
Preparation of lessons	20,00
Preparation for assessment activities	27,50
Resolution of case studies	7,50
Total hours	**90,00**

TEACHING METHODOLOGY

Theoretical activities. Lectures on the subject by the teacher with the participation of the students in the solution of specific questions.

Practicals and seminars. Learning through the resolution of exercises, problems and tasks, as well as the preparation of reports. These activities will be carried out individually or collectively in small groups in the computer classroom (practical sessions) or in the theory class (seminars).

EVALUATION

The evaluation of the subject will be based on three elements:

- Practical work. Individual assignments collected in the practical classes. Percentage of the total mark: 15%.
- Seminars. A group task on a subject related to the subject. Percentage of the total mark: 15%.
- Examination. Percentage of the total mark: 70%.

In order to pass the course it will be necessary, but not sufficient, that the student obtains a minimum mark of 4 in the exam, scored out of 10, and a minimum mark in the joint evaluation of the practicals and seminars, jointly scored out of 3.

Those students who do not pass the minimum grade in the practical and seminars will have to sit an exam on both subjects in the second general exam.

Students who have a minimum mark in practicals and seminars, but not in the exam, will be able to keep their mark in the second session.

REFERENCES

- S. N. Wood. (2017). Generalized Additive Models. An Introduction with R (Second Edition). CRC Press. Taylos & Francis Group.

Complementary references:

- Faraway, J. J. (2016). Extending the Linear Model with R. Taylor & Francis, 2016.
- G. James, D. Witten, T. Hastie, R. Tibshirani (2021). An Introduction to Statistical Learning with Applications in R (Second Ediction). Springer.
- T. Hastie, R. Tibshirani and J. Friedman (2021). The Elements of Statistical Learning. Data Mining, Inference, and Prediction (Second Ediction). Springer