Universitat Autònoma de Barcelona

2022/2023

Statistical Analysis

Code: 104364
ECTS Credits: 6

Degree	Type	Year	Semester
2503758 Data Engineering	OB	3	1

Contact

Name: Walter Andrés Ortiz Vargas
Email: walterandres.ortiz@uab.cat

Use of Languages

Principal working language: spanish (spa)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: Yes

Prerequisites

It is recommended to have taken the subject of probabilistic and statistical descriptions; notions of algebra and calculus.

Objectives and Contextualisation

The course to introduce the student to learning and stattiscal analysis using the following techniques:

1.Basic topics of statistical inference and analysis of variance of a pathway.

2.Basic elements involved in statistical analysis of the simple and multilinear regression models.

3.Tools involved in statistical analysis (estimation of paramters, descomposition of variability , ANOVA table, contrasts) of the desings by complete and non complete randomized blocks.

4.The subject also aims to familiarize students with the use of software R.

Competences

- Analyse data efficiently for the development of smart systems with the capacity for autonomous learning and/or data mining.
- Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
- Search, select and manage information and knowledge responsibly.
- 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.
- Use techniques of probability and statistics to analyse and model complex phenomena and solve optimisation problems.

Learning Outcomes

1. Correctly interpret the result of a test or statistical model for the population analysis of experimental data or the validation of an algorithm.

2. Design and implement an integrated strategy of statistical techniques and artificial intelligence for the development of descriptive and predictive systems.
3. Develop critical thinking and reasoning and know how to communicate it effectively in both your own language and in English.
4. Search, select and manage information and knowledge responsibly.
5. 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.

Content

1. Elements of statistical inference

1.1 Population and sample, parameters and statistics

1.2 Probability and inference distributions

1.3 Punctual and interval estimation

1.4 Hypothesis testing basics

1.4.1 Approach of a statistical hypothesis

1.4.2 Test for the mean

1.4.3 Test for variance

1.4.5. Three equivalent rejection or acceptance criteria

1.4.6. Hypothesis for two means: comparison of two treatments

1.4.7. Test for equality of variances

1.4.8. Paired populations (comparison of two averages with dependent samples)

2. Principles of experimental design

2.1 The design of experiments today

2.2 Basic definitions in the design of experiments

2.3 Stages in the design of experiments

2.4 Practical considerations on the use of statistical methods

2.5 Classification and selection of experimental designs

3. Analysis of Variance

3.1 Completely random design and ANOVA

3.2 Multiple range comparisons or tests

3.3 Verification of the model assumptions

3.4 Choice of sample size

3.5 Nonparametric methods in the analysis of Variance

3.5.1 Kruskal Wallis test

4. Simple and multiple linear regression

4.1 Simple linear regression

4.1.1. Hypothesis testing in simple linear regression

4.1.2. Quality of the adjustment in simple linear regression

4.1.3. Estimation and prediction by interval in simple regression (Least squares method. Maximum likelihood estimators.)

4.2 Multiple linear regression

4.2.1. Hypothesis testing in multiple linear regression

4.2.1. Confidence and prediction intervals in multiple regression

4.2.3. Matrix expression of the model and the estimators of the coefficients. Interpretation of the coefficients of the multiple model.

4.2.4 "Binding" test to resolve linear constraints on coefficients

5 Generalized and mixed linear models.

5.1 Randomized designs for complete and incomplete blocks

5.2 Design of complete blocks at random

Methodology

The subject is structured based on theory classes, problems and practices. Theory classes will introduce the concepts and techniques described in the course syllabus. you can continue to use the recommended basic bibliography. Problem classes aim to work on and understand statistical concepts. The lists of problems and, once they have been solved in class, also the solutions will be posted on the Virtual Campus. The aim of the practices is the use of statistical software R, to obtain and clarify the results of the procedures that have been introduced in the classes of theory and problems. The statement of each practice will be posted on the Virtual Campus in advance.

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.

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Practice classes	14	0.56	4, 3, 2, 1, 5
Problem Classes	14	0.56	4, 3, 2, 1, 5
Theory Classes	26	1.04	4, 3, 2, 1, 5
Type: Autonomous			
Studi	60	2.4	4, 3, 2, 1, 5

Assessment

The assessment of the subject will consist of:

1. Midterm exam (30%)

2. Final exam (40%)

3. Exam of problems (15%)

4. Continuous assessment and practices (15%)

None of the Assessment activities removes material for the final exam. The final grade will be the weighted average of the activities. No minimum grade policy is set for any activity. If applying the weights mentioned above the student's grade is 5 or higher, the subject is considered passed and this may not be the subject of a new assessment. A student is considered to be "Not assessed" in the subject as long as there is no participation in any of the assessment activities. Therefore, it is considered that a student who performs some component of Continuous Assessment can no longer qualify for an "Unassessed".

Recovery Process "To participate in the Recovery Process, students must have previously been assessed in a set of activities that represent a minimum of two thirds of the total grade of the subject or module." Section 3 of Article 112b. Recovery (UAB Academic Regulations). Students must have obtained an average grade of the subject between 4.0 and 4.9. The date of this test will be scheduled in the calendar of examinations of the Faculty. The student who presents and passes it will pass the subject with a grade of 5. Otherwise he will keep the same grade.

Irregularities in the Evaluation

Without prejudice to Other disciplinary measures deemed appropriate, and in ACCORDANCE with current academic regulations, "in the event that the student commits any Irregularities that may lead to a significant variation in the rating of an assessment actor, he / she will be graded with an 0 this actor of evaluation, independently of the disciplinary Process that can instruct. in case that produce Several Irregularities in the Acts of evaluation a same subject, the final qualification of this subject will be 0 ". Section 10 of Article 116. Results of the Evaluation. (UAB Academic Regulations) The Proposed Evaluation may undergo some modification depending on the restrictions on attendance imposed by the health authorities.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Continuous evaluation and practices	15%	30	1.2	4, 3, 2, 1, 5
Final exam	40%	2	0.08	4, 3, 2, 1, 5
Midterm exam	30%	2	0.08	4, 3, 2, 1, 5
Problem assessment	15%	2	0.08	4, 3, 2, 1, 5

Bibliography

1. Brenton, R. C. (2008). Linear models: the theory and applications of analysis of variance. ISBN: 978-0-470-0566-6.

2. Arnold, E. y Davis, Ch. S. (2002). Statistical methods for the analysis of repeated measurements. Springer.

3. Dobson, A.J. y Barnett, A.G. (2008). An introduction to generalized linear models. Series: Chapman & Hall/CRC texts in Statistical Science.

4. Fisher, R.A. (2003). Statistical methods, experimental design, and scientific inference. ISBN: 978-0-19-852229-4.

5. Gutiérrez P.H. (2003). Análisis y diseño de experimentos. McGraw-Hill.

6. Hocking, R. R. (2003). Methods and applications of linear models: regression and the analysis of variance. Wiley Series in Probability and Statistics. ISBN: 978-0-471- 23222-3.

7. Kish, L. (2004). Statistical design for research. Wiley Interscience.

8. Lindman, H. R. (1992). Analysis of variance in experimental design. Springer-Verlag.

9. Kuehl, R. O. (2001). Diseño de experimentos. Principios estadísticos del diseño y análisis de investigación. Thomson Learning.

10. Peña, D. (2002). Regresión y diseño de experimentos. Alianza.

11. Montgomery, D. C. (2002). Diseño y análisis de experimentos. Limusa-Wiley.

12. Scheiner, S.M. (2001). Design and analysis of ecological experiments. Oxford University Press.

13. Toutenburg, H. (2002). Statistical analysis of designed experiments. Springer.

Software

R-COMANDER, R-STUDIO.