Universitat Autònoma de Barcelona

**2021/2022**

# Complex Data Modelling

Code: 104864
ECTS Credits: 6

| Degree | Type | Year | Semester |
|---|---|---|---|
| 2503852 Applied Statistics | OB | 3 | 2 |

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: Joan del Castillo Franquet
Email: Joan.DelCastillo@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

## Teachers

Rosario Delgado de la Torre

## Prerequisites

It is recommended that you have passed the Calculus, Probability and Inference subjects. A minimum knowledge of Excel and R software is needed.

## Objectives and Contextualisation

Knowing tools to evaluate and quantify risk: theory of extreme values and Bayesian networks.

## 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.
- 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 other models for complex data (functional data, recount data etc.).
3. Critically assess the work done on the basis of quality criteria.
4. Establish the experimental hypotheses of modelling.
5. Identify the stages in problems of modelling.
6. Identify the statistical assumptions associated with each advanced procedure.
7. Make effective use of references and electronic resources to obtain information.
8. Make slight modifications to existing software if required by the statistical model proposed.
9. Prepare technical reports within the area of statistical modelling.
10. Reappraise one's own ideas and those of others through rigorous, critical reflection.
11. 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.
12. 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.
13. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
14. Use graphics to display the fit and applicability of the model.
15. Validate the models used through suitable inference techniques.

## Content

The subject is structured in two parts:
Topic-1: Evaluation of risk with Bayesian networks.
Introduction. From the Bayes Formula to the Bayesian Networks.
Inference with Bayesian networks.
Parameter and structure learning.
Bayesian classifiers as a tool for risk assessment.

Topic 2: Complex Systems and extreme values.
Introduction to Statistical Modeling.
Complex Systems.
Distributions on a threshold. Selection and diagnosis.
Classical theory of extreme values.

*Unless the requirements enforced by the health authorities demand a prioritization or reduction of these contents.*

## Methodology

The subject is structured from theoretical classes, problems and practices. The follow-up of the subject must be present, but it will be necessary to extend the teacher's explanations with the student's autonomous study, with the support of the reference bibliography.
The class of problems will be devoted to the resolution oriented to some problems proposed. Students' participation in the problem classes will be especially valued. Practical classes will introduce Excel and R software tools with statistical applications. You will have to deliver some practical work.

*The proposed teaching methodology may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.*

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 | | | |
| Practices (deliveries, controls) | 12 | 0.48 | 1, 3, 9, 8 |
| Problems | 14 | 0.56 | 2, 4, 14, 5, 6, 13, 15 |
| Theory | 26 | 1.04 | 2, 1, 10, 3, 9, 4, 14, 5, 6, 8, 13, 11, 12, 7, 15 |
| Type: Supervised | | | |
| Tutorials | 10 | 0.4 | 10, 3, 11, 12, 7 |
| Type: Autonomous | | | |
| Practical work with computer tools | 30 | 1.2 | 1, 3, 9, 8, 15 |
| Study and think problems | 40 | 1.6 | 4, 14, 5, 6, 13, 15 |

## Assessment

The final qualification of this subject is obtained as the average of the qualifications of the two parts of the syllabus (exposed in the Contents). The parties will be assessed with deliveries of exercises, controls of problems, practices, and exams. The examinations can only be recovered together at the end, as long as the student has previously achieved a 3.5 in each partial. Work in general is not recovered.

*Student's assessment may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.*

## Assessment Activities

| Title | Weighting | Hours | ECTS | Learning Outcomes |
|---|---|---|---|---|
| Deliveries and controls Unit-1 | 17% | 7 | 0.28 | 2, 1, 10, 3, 9, 4, 14, 5, 6, 8, 13, 11, 12, 7, 15 |
| Deliveries and controls Unit-2 | 17% | 7 | 0.28 | 2, 1, 10, 3, 9, 4, 14, 5, 6, 8, 13, 11, 12, 7, 15 |
| Final exam Theme-1 | 33% | 2 | 0.08 | 2, 1, 10, 3, 9, 4, 14, 5, 6, 8, 13, 11, 12, 7, 15 |
| Final exam Theme-2 | 33% | 2 | 0.08 | 2, 1, 10, 3, 9, 4, 14, 5, 6, 8, 13, 11, 12, 7, 15 |

## Bibliography


- Norman Fenton and Martin Neil, "Risk Assessment and Decision Analysis with Bayesian Networks", CRC Press. A Chapman & Hall Book, 2013.
- McNeil, A. J., Frey, R. and Embrechts P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press
- R. Nagarajan, M. Scutari and S. Lèbre, "Bayesian Networks in R with applications in Systems Biology", Springer, 2013.
- Cole, S. (2001). An introduction to statistical modeling of extreme values. Springer. London.

## Software

Excel and R.