



# Course guide

## 270411 - ME - Statistical Modeling

Last modified: 13/07/2023

**Unit in charge:** Barcelona School of Informatics  
**Teaching unit:** 715 - EIO - Department of Statistics and Operations Research.  
**Degree:** BACHELOR'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2021). (Compulsory subject).  
**Academic year:** 2023    **ECTS Credits:** 6.0    **Languages:** Catalan

### LECTURER

---

**Coordinating lecturer:** CARINA GIBERT OLIVERAS - SERGI RAMIREZ MITJANS

**Others:** Primer quadrimestre:  
CARINA GIBERT OLIVERAS - 11, 12  
SERGI RAMIREZ MITJANS - 11, 12

### PRIOR SKILLS

---

Introduction to Statistics  
Probability theory  
statistical inference  
simple statistical models  
data visualization  
basic programming  
R basic skills  
Algebra

### REQUIREMENTS

---

- Prerequisite IE-GIA

### DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

---

**Specific:**

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

**General:**

CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CG8. Perform an ethical exercise of the profession in all its facets, applying ethical criteria in the design of systems, algorithms, experiments, use of data, in accordance with the ethical systems recommended by national and international organizations, with special emphasis on security, robustness, privacy, transparency, traceability, prevention of bias (race, gender, religion, territory, etc.) and respect for human rights.

**Transversal:**

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.

**Basic:**

CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

CB4. That the students can transmit information, ideas, problems and solutions to a specialized and non-specialized public.

**TEACHING METHODOLOGY**

---

The course consists of two hours of theory and two laboratory hours per week. In theory classes, the inverted class scheme will be practiced whenever possible. bring preparations before each class. The master class scheme will be used on a one-off basis when the teacher needs to clarify complex concepts that have not become clear with the materials previously distributed in class. The theory class will be mainly dedicated to the presentation of cases and the development of interactive activities with students such as the discussion of cases, the development of problems or the completion of short questionnaires. Students will perform in large groups a practical work with data that they will look for themselves and that will fulfill certain characteristics set by the teaching staff. With this data each team will carry out the practice sessions, each week applying the techniques of the topic worked on in the theory session. The teacher will monitor all the work teams on a weekly basis in the laboratory sessions. At the end of the course, the teams will present their results in a pooling session where all the projects will be discussed together.

**LEARNING OBJECTIVES OF THE SUBJECT**

---

- 1.Design solvent and goal-oriented test and training games
- 2.Identify which predictive model is appropriate for a specific problem and specific data
- 3.Construct and interpret valid models for the temporal evolution of a numerical variable
- 4.Identify classes in a data set and know how to validate and interpret them conceptually
- 5.Characterize multivariate relationships in a data set with factor analysis techniques
- 6.Be able to do basic unsupervised analysis of a textual database with basic techniques of topic modeling and multivariate analysis by textual data
- 7.Know how to build and validate the right model for a new real situation
- 8.Know how to integrate the contents of the different topics of this course and the previous ones in a global solution for a complex problem
- 9.Know how to plan in the long term the modeling of a real complex problem and solve it throughout the course as a team

**STUDY LOAD**

---

Type	Hours	Percentage
Self study	90,0	60.00
Hours small group	30,0	20.00
Hours large group	30,0	20.00

**Total learning time:** 150 h



## CONTENTS

---

### Generalized linear models

**Description:**

Introduction to the concepts of generalized linear models. Logistics models

### Experimental design

**Description:**

Complete and fractional 2k designs. Sensitivity and explicability analysis of the models. Identification of main effects and interactions. Design of training sets for machine learning. Design of test sets for validation of data models

### Time series

**Description:**

Introduction to stochastic processes. Timeline vs. Time Series Box-Jenkins Methodology Main models of time series: MA, AR, ARIMA, SARIMA (concept and case study)

### Clustering

**Description:**

Introduction. Main classification models. Distances.

### Profiling

**Description:**

Description of the classifications from the study of significance of variables

### Factorial analysis

**Description:**

Dimensionality reduction methods

### Textual analysis

**Description:**

corpus preprocessing and stopwordterm document matrix ACP on this (document classification)

## ACTIVITIES

### Teamwork

#### Description:

Students are organized into groups and look for real data that meet certain requirements set by the teacher. They use them to apply the techniques and methodologies that are seen throughout the course. At the end they present a report with the results and make an oral presentation with the most relevant results of the study

#### Specific objectives:

1, 2, 3, 4, 5, 6, 7, 8, 9

#### Related competencies :

CG8. Perform an ethical exercise of the profession in all its facets, applying ethical criteria in the design of systems, algorithms, experiments, use of data, in accordance with the ethical systems recommended by national and international organizations, with special emphasis on security, robustness, privacy, transparency, traceability, prevention of bias (race, gender, religion, territory, etc.) and respect for human rights.

CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.

CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

CB4. That the students can transmit information, ideas, problems and solutions to a specialized and non-specialized public.

**Full-or-part-time:** 68h

Laboratory classes: 27h

Self study: 41h



### Quiz 1

**Description:**

During the course there will be short answer tests to fix learning pieces. It will be done at the end of certain lab classes

**Specific objectives:**

2

**Related competencies :**

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

### Quiz 2

**Description:**

During the course there will be short answer tests to fix learning pieces. It will be done at the end of certain lab classes

**Specific objectives:**

2, 3

**Related competencies :**

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.



### Quiz 3

**Description:**

During the course there will be short answer tests to fix learning pieces. It will be done at the end of certain lab classes

**Specific objectives:**

2, 3

**Related competencies :**

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.



### Initial presentation of the practice

**Description:**

Initial presentation of the practice

**Specific objectives:**

1, 2, 3, 4, 5, 6, 7, 8, 9

**Related competencies :**

CG8. Perform an ethical exercise of the profession in all its facets, applying ethical criteria in the design of systems, algorithms, experiments, use of data, in accordance with the ethical systems recommended by national and international organizations, with special emphasis on security, robustness, privacy, transparency, traceability, prevention of bias (race, gender, religion, territory, etc.) and respect for human rights.

CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.

CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

CB4. That the students can transmit information, ideas, problems and solutions to a specialized and non-specialized public.

**Full-or-part-time:** 12h

Laboratory classes: 2h

Self study: 10h

### Practical final presentation

**Description:**

Practical final presentation

**Full-or-part-time:** 10h

Self study: 10h



### Theory classes of the subject syllabus

**Description:**

Theory classes of the subject syllabus

**Specific objectives:**

2, 3, 4, 5, 6, 7

**Related competencies :**

CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.

CG4. Reasoning, analyzing reality and designing algorithms and formulations that model it. To identify problems and construct valid algorithmic or mathematical solutions, eventually new, integrating the necessary multidisciplinary knowledge, evaluating different alternatives with a critical spirit, justifying the decisions taken, interpreting and synthesizing the results in the context of the application domain and establishing methodological generalizations based on specific applications.

CE20. To select and put to use techniques of statistical modeling and data analysis, assessing the quality of the models, validating and interpreting.

CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.

CE01. To be able to solve the mathematical problems that may arise in the field of artificial intelligence. Apply knowledge from: algebra, differential and integral calculus and numerical methods; statistics and optimization.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

**Full-or-part-time:** 60h

Theory classes: 30h

Self study: 30h

## GRADING SYSTEM

(T) Teamwork done throughout the course 20%

(O) Oral knowledge control test 10% (discussion with the teacher in the oral presentation of the team work)

(WT) Quality and performance of the work team (TG). 10%

(C) Oral and written communication 10%

(E) Work team ethics and work itself 10%

(G) Gender perspective of the team and work 10%

(A) Attendance and participation in classes and laboratories (AP). 10%

(Q) 3 Quiz throughout the course 20%

$$N=0,2T+0,1*O+0,1*WT+0,1*C+0,1*E+0,1*G+0,1*A+0,2*Q$$

$$Q=(Q1+Q2+Q3)/3$$





## BIBLIOGRAPHY

---

### Basic:

- Bruce, Peter; Bruce, Andrew; Gedeck, Peter. Practical statistics for data scientists: 50+ essential concepts using R and Python. 2nd ed. O'Reilly, [2020]. ISBN 9781492072942.
- Maindonald, J. H; Braun, John. Data analysis and graphics using R : an example-based approach. 3rd ed. Cambridge University, 2010. ISBN 9780521762939.
- Hyndman, R.J.; Athanasopoulos, G. Forecasting: principles and practice. 3rd print edition. [Melbourne]: O Texts, 2021. ISBN 9780987507136.
- Nielsen, Aileen. Practical time series analysis: prediction with statistics and machine learning [on line]. Sebastopol: O'Reilly Media, Inc, 2019 [ Consultation: 26/09/2023]. Available on : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5899923>. ISBN 9781492041658.