



## Course guide

### 270406 - IE - Introduction to Statistics

Last modified: 02/02/2024

**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:** English

#### LECTURER

---

**Coordinating lecturer:** CARINA GIBERT OLIVERAS

**Others:**

#### PRIOR SKILLS

---

Coneixements matemàtics de batxillerat

#### 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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems . To be able to apply all these for solving problems.

##### Generical:

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.

## TEACHING METHODOLOGY

---

The course consists of two hours of theory and two hours of lab per week.

In theory classes (lectures), the inverted class scheme will be practiced whenever possible.

On the website of the subject, the calendar of the course will be posted and the materials to be prepared before each class. The master class scheme is covered promptly when the lecturer needs to clarify complex concepts that have not been solved within materials distributed prior to the class. The theory class will be dedicated primarily 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 realization of short specific questionnaires.

The students will be organized by groups (teams) and will work in a practical project with data that they will look for themselves and that will fulfil certain characteristics set by the professors. With these data, each team will carry out the practice sessions with the techniques of the topic studied during the theory session of each week. The lecturer will monitor weekly all the work teams during the lab sessions.

At the end of the course, the teams will present their results in a public session where all the projects will be discussed together.

## LEARNING OBJECTIVES OF THE SUBJECT

---

1. Getting familiar with the tools of basic statistics to be able to treat data correctly and internalize the statistical methodology as a basic scheme of extraction of relevant information on complex phenomena
2. To select the relevant data to support a specific question.
3. Designing the eligibility criteria of a sample correctly to answer a real problem
4. Designing basic experiments to study real problems
5. To perform basic data preprocessing
6. To select the statistical modeling methods most appropriate to the problem, in relation to the structure of the available data, the objectives of the study and the subsequent uses of the model results
7. To build the statistical models correctly from the data, making use of the necessary software, the context of the reference problem and present it publicly
8. To apply in an integrated way the statistical knowledge obtained from the classes into the analysis of a real data set (taking advantage of open data sources) responding to a reference problem of any real area relevant to artificial intelligence, such as health, environment, sustainability, industry4.0
9. To perform and develop practical works and projects with a gender perspective
10. Integrating teamwork mechanisms in carrying out practical works
11. Deftly deal with the computer tools necessary to solve the real problems proposed with the basic statistical techniques seen during the course.
12. To Interpret and contextualize the statistical models built from data
13. Incorporating the ethical recommendations of the EC regarding AI to practical work
14. To validate the models obtained and make a critical interpretation of the results from a technical point of view and contextualize the results within the framework of the problem.
15. To carry out an automatic report with the descriptive analysis of a Database, the validated models, and the integrated and critical analysis of the results in relation to the context of the reference problem.
16. Publicly present a statistical report that includes descriptive analysis, models and conclusions, adequately communicated in technical audiences and / or without technical competences

## STUDY LOAD

---

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

**Total learning time:** 150 h



## CONTENTS

---

### Descriptive Analysis of Data

**Description:**

We will work on how to use numerical and graphical statistical tools to describe a set of data, as well as the automatic reporting tools necessary to perform automatic reporting with this description.

### Introduction to probability theory

**Description:**

The basic notions of probability will be provided to understand the concept of uncertainty and the main probabilistic formalisms to model it, including concepts of conditioned probability and the Bayes theorem, relevant in later subjects.

### Introduction to sampling theory

**Description:**

simple random sample concept, sampling theory, sampling type

### Statistical Inference

**Description:**

hypothesis testing, p-value concept, confidence intervals. Limitations in real applications of classical inference. Nonparametric inference, Fisher permutation test. Hypothesis tests in statistical learning

### Introduction to experimental design

**Description:**

differences between sample and experimental studies. The design of experiments in software validation. Biases and scalability

### Regression

**Description:**

Basic model (simple linear regression, mean least squares). Goodness of fit measures, validation. Multiple linear regression. General linear model (ANOVA, ANCOVA)

## ACTIVITIES

### Teamworking

**Description:**

Students are organized into groups and look for real data sets that match certain requirements defined by the professor. Data are used to apply the techniques and methodologies that are seen throughout the course. At the end of the course, they will present a report with the results and make an oral presentation with the most relevant results of the final project.

**Specific objectives:**

8, 10

**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.

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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems . To be able to apply all these for solving problems.

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:** 77h

Laboratory classes: 27h

Self study: 50h

### quiz 1

**Description:**

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

**Specific objectives:**

1, 2, 3, 4, 5, 6, 9, 11

**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.

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.

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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems . To be able to apply all these for solving problems.

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.

**Full-or-part-time:** 0h 30m

Guided activities: 0h 30m



## Quiz 2

### Description:

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

### Specific objectives:

6, 7, 8, 10, 12, 14

### 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.

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.

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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems . To be able to apply all these for solving problems.

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.

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.

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.

**Full-or-part-time:** 0h 30m

Guided activities: 0h 30m



### Initial practical work presentation

**Description:**

Initial presentation of the practical work

**Specific objectives:**

9, 13, 15

**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.

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.

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.

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.

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.

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

Laboratory classes: 2h

Self study: 10h



### Final practical work presentation

**Description:**

Final practical work presentation

**Specific objectives:**

7, 8, 9, 10, 11, 12, 13, 14, 15, 16

**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.

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.

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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems. To be able to apply all these for solving problems.

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.

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.

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.

**Full-or-part-time:** 3h

Guided activities: 3h

### Lectures

**Description:**

Lectures

**Specific objectives:**

1

**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.

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.

CE02. To master the basic concepts of discrete mathematics, logic, algorithmic and computational complexity, and its application to the automatic processing of information through computer systems. To be able to apply all these for solving problems.

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

Theory classes: 30h

Self study: 30h

## GRADING SYSTEM

---

- (T) Teamwork carried out during the course 20%
- (O) Oral knowledge control test 10% (discussion with teachers in the oral presentation of teamworks)
- (CT4) Quality and performance of the work team (TG). 10%
- (CT3) Oral and written communication 10%
- (CT8) Team and work gender perspective 10%
- (E) Ethics of the work team and the work itself 10%
- (A) Attendance and participation in classes and laboratories (AP). 10%
- (Q) 2 Quiz throughout the course 20%

$$N = 0,2 * T + 0,1 * O + 0,1 * CT4 + 0,1 * CT3 + 0,1 * E + 0,1 * CT8 + 0,1 * A + 0,2 * Q$$

$$Q = 0,5 * Q1 + 0,5 * Q2$$

## BIBLIOGRAPHY

---

### Basic:

- Agresti, Alan; Franklin, Christine. Statistics: the art and science of learning from data. 4th ed. Pearson Education, 2017. ISBN 9781292164878.
- Bruce, Peter; Bruce, Andrew; Gedeck, Peter. Practical statistics for data scientists : 50+ essential concepts using R and Python. 2nd ed. Sebastopol, California: O'Reilly Media, Inc, 2020. ISBN 9781492072942.
- Dekking, F.M; Kraaikamp, C.; Lopuhaä, H.P. A modern introduction to probability and statistics: understanding why and how. London: Springer, 2005. ISBN 9781846281686.
- Maindonald, John; Braun, John. Data analysis and graphics using R : an example-based approach. 3rd ed. Cambridge: Cambridge University, 2010. ISBN 9780521762939.
- Moore, D.S.; McCabe, G.P.; Craig, B.A. Introduction to the practice of statistics. 10th ed. New York: WH Freeman, 2021. ISBN 9781319383664.