

# STATISTICS & DATA ANALYSIS

**IE University**

Professor: **MARCO CASERTA**

E-mail: [mcaserta@faculty.ie.edu](mailto:mcaserta@faculty.ie.edu)

Degree course: SECOND

Semester: 1º

Category: BASIC

Number of credits: 6.0

Language: English

## PREREQUISITES

Students must be familiar with the mathematical concepts presented in the first academic year.

## SUBJECT DESCRIPTION

Statistics uses mathematical tools to organize and summarize data obtained from the real world, and to draw conclusions derived from a correct interpretation of these data. In the business world, statistics can help assess the attractiveness of a business opportunity, increase customer satisfaction, choose between different investment possibilities, analyze and improve production processes, etc.

Students following this course will learn how to define the data required in different situations characterized by uncertainty, how to collect and summarize these data, and how to make decisions based on data analysis. This course also provides the theoretical and practical bases for other courses in the degree, such as Marketing Research and Business Decision Making.

## OBJECTIVES AND SKILLS

The objective of this course is to provide students with the tools to organize and understand data and to make use of this information in business applications. At the end of the course you should be able to:

- Use of python and some statistical libraries to import, manipulate, plot, and analyze data.
- Describe data by means of graphs or numbers, and understand in which contexts each of these descriptive tools is useful.
- Understand patterns of randomness that can affect business activities and relate them to known probability distributions.
- Understand the differences between population and sample distributions.
- Derive confidence intervals for a parameter.
- Make inferences by understanding the concept of null and alternative hypotheses and interpret outputs of hypothesis testing.
- Test for differences between populations.
- Use statistical methods for decision-making in a business context.

Additionally, the course will focus on the acquisition or reinforcement of generic skills:

- The ability to summarize and present information in a meaningful way.
- The ability to build an abstract model to address an economic problem.
- The ability to quickly identify the tools that are useful in business situations.

## METHODOLOGY

The course is articulated on a mix of lectures and computer-based sessions. During the lectures, we will first motivate the need of theoretical tools by means of a real-world example. Next, the relevant theory will be presented. During the lab sessions, the students will be asked to address a challenging real-world problem and to employ the tools developed in class to analyze the problem, model it, attain a solution, and interpret it in the context of the underlying real-world problem.

Before all sessions, you are required to read the textbook sections as indicated in the syllabus. Reading the textbook in advance will allow you to get the most out of each lecture. When reading the textbook portions prior to each lecture, you must look at the examples but you do not need to solve them.

You will have to prepare a number of exercises before most sessions, and be ready to discuss them during the class. Participation will be graded. At the beginning of some sessions, a brief quiz covering some aspects of the reading and/or exercises will be given. Marks obtained on these quizzes will be included in the final grade.

Teaching methodology	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	20.0 %	30 hours
Discussions	20.0 %	30 hours
Exercises	14.67 %	22 hours
Group work	30.0 %	45 hours
Other individual studying	15.33 %	23 hours
TOTAL	100.0 %	150 hours

## EVALUATION CRITERIA

Criteria	Percentage	Comments
Class Participation	15 %	See comments below on participation.
Group Project	20 %	Group project.
Quizzes and Other Activities	30 %	Short paper-based tests.
Final Exam	35 %	Omnicomprehensive final exam.

### A. Class participation and discussion

Class participation will be evaluated based on the following criteria:

- Quality (not quantity) of your participation in class discussion: The most important dimension of participation concerns what it is that you are saying. A high quality comment reveals depth of insight, rigorous use of case evidence, consistency of argument, and realism.
- Frequency refers to the attainment of a threshold quantity of contributions that is sufficient for making a reliable assessment of comment quality. The logic is simple: if contributions are too few, one cannot reliably assess the quality of your remarks. However, once threshold quantity has been achieved, simply increasing the number of times you talk does not automatically improve your evaluation. Beyond the threshold, it is the quality of your comments that must improve. In particular, one must be especially careful that in claiming more than a fair share of “airtime”, quality is not sacrificed for quantity.
- Finally, your attempts at participation should not be such that the instructor has to “go looking for you”. You should be attempting to get into the debate on a regular basis.

You might want to avoid being classified as one of the following types of students:

- Repeaters, i.e., students that, consciously or unconsciously, make comments that are really just repeats/rephrasing of what has already been said (by other students, or you). This wastes time and adds nothing to learning.
- Ramblers, i.e., students that take a lot of time to say simple things or they may tell long personal/professional stories, or they roam into topics that are not relevant, or simply make low-quality comments just to participate. They waste valuable time and prevent other students from being able to participate.
- Students that have been distracted (by Facebook, etc.) or who have stopped paying attention and then, later on, when they realized they have missed a term or concept, they ask you about it.

During some of the asynchronous classes, videos and/or simulation activities with review questions will be assigned. These activities will be evaluated along multiple dimensions (participation, quality of the answers, etc.) and the grade will contribute to the final participation grade.

## **B. Group report and presentation**

Each group should be composed of 4 to 5 students and must prepare a group report due at the end of the course (more details about intermediate and final deadline will be periodically provided during the course.) The group project will consist in the identification of real-world dataset, publicly available on the web, which should be used to identify and analyze interesting and relevant questions. Answers should be provided using the tools seen in class. Specifically, the use of python and its libraries to analyze the dataset is mandatory.

Every submission will be delivered using turnitin following the appropriate link provided on campus online. At the end of the semester, you must submit the full report including all sections. The final version should include edited versions of the previously submitted sections following the recommendations of your professor.

Make sure the report is easy to read. Consider using bullets, headings, etc., to make it easy to follow. Avoid being too technical in the report: provide a fact-based rationale for your comments but make sure that your explanations and recommendations are understandable to someone with very little statistical knowledge.

Since you will be using publicly available datasets, which implies that other people might have worked on questions similar to yours, make sure that you fully understand the **IE policy on plagiarism and cheating**. If you have any doubts about whether a certain course of action is appropriate, assume it is not and avoid it.

## **C. Quizzes and Other Activities**

At the beginning of some sessions, you will be given a short quiz based on required readings and exercises for the session. Quizzes will be announced in advance and can be both paper and computer-based. The final grade for this item will be computed excluding the lowest grade. In addition, during some of the lab sessions, short exercises based on the use of collaborative notebooks will be assigned and evaluated.

## **D. Final exams**

There will be one final exam, scheduled on the last session of the course. The final exam will be computer-based. Therefore, you must bring your own computer and ensure that you will be able to connect to the internet and use the resources needed to perform statistical analysis as seen in class. The exam is open book.

In order to pass the course, you need a **minimum grade of 3.5 in the final exam**. If your grade in the final exam does not reach the threshold value of 3.5, you will fail the course, even in the case in which your weighted average (computed using the table above) exceeds 5.0.

## **PROFESSOR BIO**

Professor: **MARCO CASERTA**

E-mail: [mcaserta@faculty.ie.edu](mailto:mcaserta@faculty.ie.edu)

Marco Caserta received his Ph.D. in Industrial Engineering and Operations Research from the University of Illinois (USA), after earning a MSc in Management Engineering from the Politecnico di Milano (Italy). He is currently a professor at IE University as well as an associate professor at the IE Business School. He teaches optimization related courses to graduate students within the International MBA and the PhD programs.

His main research interest is concentrated on the design and development of metaheuristic-based algorithms for very large scale real-world optimization problems, with a special focus on logistics, telecommunication and transportation related problems. He has published a number of papers in journals in the area of operations research/management science.

## **OTHER INFORMATION**

### **OFFICE HOURS; CONTACT INFORMATION**

Office hours: Right after the designated classes (upon appointment)

Contact details:

E-mail: [mcaserta@faculty.ie.edu](mailto:mcaserta@faculty.ie.edu)

