STATISTICS & DATA ANALYSIS

Grado en Administración de Empresas / Bachelor in Business Administration BBA SEP-2023 STA-NBA.2.M.D

Area Operations and Business Analytics

Number of sessions: 35

Academic year: 23-24

Degree course: SECOND

Number of credits: 6.0

Semester: 2º

Category: BASIC

Language: English

Professor: **MAUD PINDARD**

E-mail: mpindard@faculty.ie.edu

Maud Pindard-Lejarraga holds a Master and Ph.D. in Business Administration and Quantitative Methods from Universidad Carlos III de Madrid, and Bachelor Degrees in Econometrics from Université Paris 1 - ENS Ulm and in Hispanic Studies from Université Paris 3 (France). Before joining IE in 2010, she worked as Teaching Assistant at Universidad Carlos III, and as Invited Researcher at HEC Paris and Rotman School of Business, University of Toronto (Canada). She also worked as an advisor in the Economic Service of the French Embassy in Spain.

Her research focuses on the macro-level evolution of industries, and on entrepreneurs' entry decisions. Her recent articles have been published in *Academy of Management Learning and Education* and the *Journal of Small Business Management*.

Office Hours

Office hours will be on request. Please contact at:

maud.pindard@ie.edu

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 Data Analysis for Economics.

LEARNING OBJECTIVES

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 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.

TEACHING METHODOLOGY

IE University teaching method is defined by its collaborative, active, and applied nature. Students actively participate in the whole process to build their knowledge and sharpen their skills. Professor's main role is to lead and guide students to achieve the learning objectives of the course. This is done by engaging in a diverse range of teaching techniques and different types of learning activities such as the following:

Learning Activity	**Weighting**	**Estimated time a student should dedicate to prepare for and participate in**
Lectures	20.0 %	30.0 hours
Discussions	20.0 %	30.0 hours
Exercises in class, Asynchronous sessions, Field Work	14.67 %	22.0 hours
Group work	30.0 %	45.0 hours
Individual studying	15.33 %	23.0 hours
TOTAL	100.0 %	150.0 hours

PROGRAM

SESSION 1 (LIVE IN-PERSON)

Plenary Sessions

- Data Collection Principles, Sampling
- Observational Studies vs Experiments
- Types of Variables
- Examining Numerical and Categorical Data)
    - Charts: Histograms, Scatter Plots, Boxplots
    - Summary Statistics: Mean, Median, Variance, Standard Deviation, Skewness, Kurtosis, Quartiles, Percentiles, Covariance, Correlation, Contingency Tables

SESSION 2 (LIVE IN-PERSON)

Plenary Sessions

- Data Collection Principles, Sampling
- Observational Studies vs Experiments
- Types of Variables
- Examining Numerical and Categorical Data)
    - Charts: Histograms, Scatter Plots, Boxplots
    - Summary Statistics: Mean, Median, Variance, Standard Deviation, Skewness, Kurtosis, Quartiles, Percentiles, Covariance, Correlation, Contingency Tables

SESSION 3 (LIVE IN-PERSON)

Plenary Sessions

- Data Collection Principles, Sampling
- Observational Studies vs Experiments
- Types of Variables
- Examining Numerical and Categorical Data)
    - Charts: Histograms, Scatter Plots, Boxplots
    - Summary Statistics: Mean, Median, Variance, Standard Deviation, Skewness, Kurtosis, Quartiles, Percentiles, Covariance, Correlation, Contingency Tables

SESSION 4 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Introduction to Python Notebooks and Colaboratory.
- Features.
- Markdown Guide.
- External Data Input and Output: Connection with Google Drive.
- Introduction to Pandas Dataframes (and Numpy).
- Descriptive Statistics with Matplotlib

    - Histograms, Bar Charts
    - Scatter Plots
    - Box Plots
- Exercises

Technical note: Introduction to Python Notebooks and Colaboratory (Google Colab)
Technical note: Features (Google Colab)
Technical note: Markdown Guide (Google Colab)
Technical note: External Data Input and Output: Connection with Google Drive (Google Colab)
Technical note: Introduction to Pandas Dataframes (and Numpy) (Google Colab)

SESSION 5 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Introduction to Python Notebooks and Colaboratory.
- Features.
- Markdown Guide.
- External Data Input and Output: Connection with Google Drive.
- Introduction to Pandas Dataframes (and Numpy).
- Descriptive Statistics with Matplotlib
    - Histograms, Bar Charts
    - Scatter Plots
    - Box Plots
- Exercises

SESSION 6 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Introduction to Python Notebooks and Colaboratory.
- Features.
- Markdown Guide.
- External Data Input and Output: Connection with Google Drive.
- Introduction to Pandas Dataframes (and Numpy).
- Descriptive Statistics with Matplotlib
    - Histograms, Bar Charts
    - Scatter Plots
    - Box Plots
- Exercises

SESSION 7 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Introduction to Python Notebooks and Colaboratory.
- Features.
- Markdown Guide.
- External Data Input and Output: Connection with Google Drive.

- Introduction to Pandas Dataframes (and Numpy).
- Descriptive Statistics with Matplotlib
    - Histograms, Bar Charts
    - Scatter Plots
    - Box Plots
- Exercises

SESSION 8 (LIVE IN-PERSON)

Plenary Sessions

- Exercises

SESSION 9 (LIVE IN-PERSON)

Plenary Sessions

- Introduction to distribution functions:
    - Discrete: Binomial, Poisson
    - Continuous: Uniform, Normal, Exponential

SESSION 10 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

Distribution functions:

Discrete: Binomial, Poisson

Continuous: Uniform, Normal, Exponential

SESSION 11 (LIVE IN-PERSON)

Plenary Sessions

- Normal distribution with python
- Exercises on the Normal Distribution

SESSION 12 (LIVE IN-PERSON)

Plenary Sessions

- Normal distribution with python
- Exercises on the Normal Distribution

SESSION 13 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

Scipy Stats Library and Distribution Functions

Cumulative distribution functions

Inverse of distribution functions

How to create a plot of a distribution function

- Exercises

SESSION 14 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

Scipy Stats Library and Distribution Functions

- Cumulative distribution functions
- Inverse of distribution functions
- How to create a plot of a distribution function

- Exercises

SESSION 15 (LIVE IN-PERSON)

Plenary Sessions

- The Central Limit Theorem
- Confidence Interval for a Population Parameter

SESSION 16 (LIVE IN-PERSON)

Plenary Sessions

- The Central Limit Theorem
- Confidence Interval for a Population Parameter

SESSION 17 (LIVE IN-PERSON)

Plenary Sessions

- The Central Limit Theorem
- Confidence Interval for a Population Parameter

SESSION 18 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Confidence Intervals with Statsmodels: Introduction and Exercises

SESSION 19 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Confidence Intervals with Statsmodels: Introduction and Exercises

SESSION 20 (LIVE IN-PERSON)

Plenary Sessions

- Exercises

SESSION 21 (LIVE IN-PERSON)

Plenary Sessions

- Introduction to Hypothesis Testing One Sample Means and Proportions
- Hypothesis Testing One Sample Means)
    - Large Samples (Variance Known)

    - Small Samples (Variance Unknown): The t distribution
- Hypothesis Testing One Sample Proportions
- Hypothesis Testing: Types of Errors

SESSION 22 (LIVE IN-PERSON)

Plenary Sessions

- Introduction to Hypothesis Testing One Sample Means and Proportions
- Hypothesis Testing One Sample Means)
    - Large Samples (Variance Known)
    - Small Samples (Variance Unknown): The t distribution
- Hypothesis Testing One Sample Proportions
- Hypothesis Testing: Types of Errors

SESSION 23 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Hypothesis Testing with Python: One sample
- Hypothesis Testing: Two samples
- Difference between two-pair and two-independent-sample tests
- Exercises

SESSION 24 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Hypothesis Testing with Python: One sample
- Hypothesis Testing: Two samples
- Difference between two-pair and two-independent-sample tests
- Exercises

SESSION 25 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Hypothesis Testing with Python: One sample
- Hypothesis Testing: Two samples
- Difference between two-pair and two-independent-sample tests
- Exercises

SESSION 26 (LIVE IN-PERSON)

Plenary Session

- Two-paired Sample Tests
- Two-independent Sample Tests Difference of Population Means
- Difference of Population Proportions

SESSION 27 (LIVE IN-PERSON)

Plenary Session

- Two-paired Sample Tests
- Two-independent Sample Tests Difference of Population Means
- Difference of Population Proportions

SESSION 28 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Two sample tests with statsmodels
- Exercises

SESSION 29 (LIVE IN-PERSON)

Plenary Session

- Case/Exercises on hypothesis testing two-samples

SESSION 30 (LIVE IN-PERSON)

Plenary Sessions

- One-Way ANOVA: Assumptions, Post-hoc Analysis

SESSION 31 (LIVE IN-PERSON)

Plenary Sessions

- One-Way ANOVA: Assumptions, Post-hoc Analysis

SESSION 32 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- One-Way ANOVA and Post-hoc analysis with statsmodels (py-8.3)
- Checking ANOVA assumptions with charts
- Exercises

SESSION 33 (LIVE IN-PERSON)

Plenary Session

- Exercises on Hypothesis Testing & ANOVA

SESSION 34 (LIVE IN-PERSON)

Lab Sessions Groups 1 and 2 (groups assignment to be defined)

- Mock Exam

SESSION 35 (LIVE IN-PERSON)

Plenary Session

- **Final Exam**

EVALUATION CRITERIA

All the graded activities, including the final exam, the quizzes, etc. must be carried out without assistance. AI-based platforms, such as Open AI and similar, are prohibited during the graded activities. Giving or receiving help, via chat or in any other form, is also prohibited. In addition, all the students must be connected to the IE wifi network during the entire duration of each activity. Students failing to follow these instructions will receive a grade of zero in the activity, and they will be reported to the BBA office and to the ethical committee when it applies.

criteria	percentage	Learning Objectives	Comments
Class Participation	15 %		See comments below on participation.
Group Project	25 %		Group project.
Activities and Notebooks	20 %		Short exercises.
Final Exam	40 %		Omnicomprehensive final exam.

RE-SIT / RE-TAKE POLICY

EVALUATION CRITERIA FOR THE FIRST ORDINARY CALL

The minimum passing grade is 5.0.

A. Class participation

Class participation will be evaluated based on your participation in class discussion, firstly, and on class attendance secondly.

B. Group project

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 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 Python.

Make sure the project is easy to read. Consider using bullets, headings, etc., to make it easy to follow. Avoid being too technical in the text: 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**.

C. Quizzes

Several quizzes and graded exercises will be done in class.

D. Final exam

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.

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 with a maximum grade of 4.9, even if your weighted average (computed using the table above) exceeds 5.0.

RE-SIT / RE-TAKE POLICY

Each student has four (4) chances to pass any given course distributed over two (2) consecutive academic years. Each academic year consists of two calls: one (1) ordinary call (during the semester when the course is taking place); and one (1) extraordinary call (or "re-sit") in June/July.

Students who do not comply with the 80% attendance requirement in each subject during the semester will automatically fail both calls (ordinary and extraordinary) for that Academic Year and have to re-take the course (i.e., re-enroll) during the next Academic Year.

The Extraordinary Call Evaluation criteria will be subject to the following rules:

- Students failing the course in the ordinary call (during the semester) will have to re-sit evaluation for the course in June / July (except those students who do not comply with the attendance rule, and therefore will not have that opportunity, since they will fail both calls and must directly re-enroll in the course during the next Academic Year).
- It is not permitted to change the format nor the date of the extraordinary call exams or deadlines under any circumstance. All extraordinary call evaluation dates will be announced in advance and must be taken into consideration before planning the summer (e.g. internships, trips, holidays, etc.)
- The June/July re-sit will consist of a comprehensive evaluation of the course. Your final grade for the course will depend on the performance in this exam or evaluation only. I.e., continuous evaluation over the semester (e.g. participation, quizzes, projects and/or other grade components over the semester) will not be taken into consideration on the extraordinary call. Students will have to achieve the minimum passing grade of 5 and the maximum grade will be capped at 8.0 (out of 10.0) – i.e., "notable" in the extraordinary call.
- Re-takers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will need to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as re-takers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained as a retaker during the ordinary call (i.e., the 3rd call) is 10.0 (out of 10.0).

After exams and other assessments are graded by the professor (on either the ordinary or extraordinary call), students will have a possibility to attend a review session (whether it be a final exam, a final project, or the final overall grade in a given course). Please be available to attend the session in order to clarify any concerns you might have regarding your grade. Your professor will inform you about the time and place of the review session.

- Students failing more than 18 ECTS credits after the June/July re-sits will be asked to leave the Program. Please, make sure to prepare yourself well for the exams in order to pass your failed subjects.
- In case you decide to skip the opportunity to re-sit for an exam or evaluation during the June/July extraordinary call, you will need to enroll in that course again for the next Academic Year as a re-taker, and pay the corresponding tuition fees. As you know, students have a total of four (4) allowed calls to pass a given subject or course, in order to remain in the program.

BIBLIOGRAPHY

Compulsory

- David Diez, Mine Çetinkaya-Rundel, Christopher Barr. (2019). *Openintro Statistics.* 4th. OpenIntro, Inc. ISBN 9781943450077 (Digital)

Free pdf available at:

https://www.openintro.org/book/os/

- Thomas Haslwanter. (2016). *An Introduction to Statistics with Python: With Applications in the Life Sciences.* 1st. Springer. ISBN 9783319803234 (Digital)

Recommended

- Newbold, Paul, Carlson, William L. & Thorne, Betty. (2013). *Statistics for Business and Economics..* 8th global edition. Pearson Prentice Hall. ISBN 9780273767060 (Printed)

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

ETHICAL POLICY

Please, check the University's Ethics Code here. The Program Director may provide further indications.