FUNDAMENTALS OF PROBABILITY AND STATISTICS

Bachelor in Data and Business Analytics BDBA SEP-2023
FPS-DBA.1.S.A

Area Operations and Business Analytics
Number of sessions: 30
Academic year: 23-24
Degree course: FIRST
Number of credits: 6.0
Semester: 1º
Category: BASIC
Language: English

Professor: **BORJA MESA**

E-mail: bmesa@faculty.ie.edu

BORJA MESA SANCHEZ

Borja Mesa-Sánchez is an expert in the area of industrial organization. After finishing his Ph.D. in Quantitative Economics (with honors) at Universidad de Alicante he joined the Department of Economics of Universidad Carlos III de Madrid as a Post-doctoral fellow. He received a Master Degree in Quantitative Economics, and his Bachelor's Degree from Universidad Carlos III de Madrid (with honors). His research is in competition policy, game theory and industrial organization. He has published in international scientific journals ranked in JCR by ISI. Borja teaches at undergraduate and graduate level and he has been the advisor for many senior thesis. He has worked in the department of training in Reuters, and he has won two literary awards.

Experience

- Visiting Professor Brown University
- Post-doctoral fellow, Universidad Carlos III de Madrid
- Adjunct Professor, Saint Louis University
- Teaching Assistant, Universidad de Alicante
- Training Assistant, Reuters

Education

- Post-doctoral fellow, Universidad Carlos III de Madrid
- Ph.D in Quantitative Economics (with honors), Universidad de Alicante Stays in European University Institute and University of York
- MSc in Quantitative Economics, Universidad de Alicante
- Bs Economics (with honors), Universidad Carlos III de Madrid

bmesa@faculty.ie.edu

SUBJECT DESCRIPTION

Do you know that there exists a positive correlation between children academic performance and parents' income? Can you think of any possible explanation? Intelligence? Better resources? No matter where you stand on this issue, you can always find statistics to back up your point of view. The truth of the matter is that there are many factors behind such a complex phenomenon, and we will not be able to disentangle such factors until we have reliable data available. What is reliable data? How important is it in representing a real-life situation? What do experts mean when they talk about representative samples? What is a sample and is there more than one way to take a sample? What insights can we draw from data? What are the numbers that matters and how can we use them to infer conclusions about our population? Answers to these questions and many more will be given in this introductory course on probability and statistics.

Statistics is the science of data, a discipline grounded in mathematics that converts raw data into actionable information. It uses mathematical tools to construct formal models, to summarize and process data and to reduce uncertainty in different environments: engineering, financial markets, the insurance industry, biomedicine, consumer behavior, presidential elections, gambling industry, physics, etc.

In this introductory course, we will focus on understanding some of the foundational concepts in statistics. Students will learn how to collect and summarize data, how to describe data graphically and numerically, understand patterns of randomness that can affect real life activities and relate them to known probability distributions. Students will also learn the concept of random variable, and the most important probability distributions, both discrete and continuous. Finally, we will go over the concept of joint probability distributions.

This course will be the basis for further subjects such as Fundamentals of Data Analysis, Probability and Statistics for Data Management and Analysis as well as Forecasting and Time Series Analysis, among others.

LEARNING OBJECTIVES

The objective of this course is to provide students with the tools to delve into data sets and to make use of this information in many different applications. At the end of the course; students should be able to:

- Describe data by means of graphs.
- Describe data using numerical measures.
- Understand in which context each of these descriptive tools is useful.
- Understand patterns of randomness that can affect real life activities and relate them to known probability distributions.
- Use the most common discrete probability distributions
- Use the most common continuous probability distributions.
- Understand the concept of joint probability distributions

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

- The ability to think analytically;
- The use of the statistical software RStudio

- The ability to think critically.

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	6.67 %	10.0 hours
Exercises in class, Asynchronous sessions, Field Work	20.0 %	30.0 hours
Group work	20.0 %	30.0 hours
Individual studying	33.33 %	50.0 hours
TOTAL	100.0 %	150.0 hours

PROGRAM

MODULE 1: DESCRIPTIVE STATISTICS

SESSIONS 1 - 2 (LIVE IN-PERSON)

Introduction to the Course: Objectives, Course Organization and Evaluation.

Topic 1. Fundamentals Elements of Statistics.

Fundamentals Elements of Statistics. Types of Data. Sampling principles and strategies. Other concepts

Readings:

Book Chapters: Chapter 1. Sections 1.1, 1.2 (NCT)

Book Chapters: Chapter 1. Sections 1.1, 1.3, 1.5, 1.6, (MBS).

Book Chapters: Chapter 1. Sections 1.1, 1.2 (OIS)

Book Chapters: Chapter 1. Section 1.1 (Devore)

SESSION 3 (LIVE IN-PERSON)

Introducing RStudio

How to install RStudio

Basic operations with RStudio

SESSION 4 (LIVE IN-PERSON)

Topic 2. Descriptive Statistics. Graphical Methods

Describing qualitative data. Frequency distribution for categorical data. Contingency tables (Cross-tables). Bar charts. Pie charts. Pareto diagrams. Time-Series plot. Frequency distributions for numerical data. Dot plots. Histograms. Stem-and-leaf display. Scatter plots. Other concepts.

Readings:

Book Chapters: Chapter 1. Sections 1.3, 1.4, 1.5, 2.4 (NCT)

Book Chapters: Chapter 2. Sections 2.1, 2.2, 2.8, 2.9 (MBS).

Book Chapters: Chapter 2. Sections 2.1, 2.2 (OIS)

Book Chapters: Chapter 1. Section 1.2 (Devore)

SESSION 5 (LIVE IN-PERSON)

Introducing RStudio

How to install RStudio

Data manipulation

SESSIONS 6 - 7 (LIVE IN-PERSON)

Topic 3. Descriptive Statistics. Numerical Measures

Summary Statistics: Mean, Median, Range, Variance, Standard Deviation, Coefficient of Variation, Chebyshev's Rule, Empirical rule, Quartiles, Percentiles, Z-score. Weighted mean, Five number summary. Box-plots. Detecting outliers. Covariance, Correlation. Other concepts.

Readings:

Book Chapters: Chapter 2. Sections 2.1, 2.2, 2.3, 2.4 (NCT)

Book Chapters: Chapter 2. Sections 2.3, 2.4, 2.5, 2.6, 2.7 (MBS).

Book Chapters: Chapter 2. Sections 2.1, 2.2 (OIS)

Book Chapters: Chapter 1. Sections 1.3, 1.4 (Devore)

MODULE 2: PROBABILITY

SESSION 8 (LIVE IN-PERSON)

RStudio

Descriptive Statistics with RStudio

SESSION 9 (LIVE IN-PERSON)

Descriptive Statistics

Topic 2. Descriptive Statistics. Graphical Methods.

Topic 3. Descriptive Statistics. Numerical Measures.

Problem Set 1

SESSIONS 10 - 11 (LIVE IN-PERSON)

Topic 4. Probability

Basic definitions. Probability and its postulates. Probability rules. Conditional probability. The Multiplicative Rule and Independent Events. Bivariate probabilities. Baye's Theorem. Sampling from small populations. Other concepts.

Book Chapters: Chapter 3. All Sections (NCT)

Book Chapters: Chapter 3. All Sections (MBS).

Book Chapters: Chapter 3. Sections 3.1, 3.2, 3.3 (OIS)

Book Chapters: Chapter 2. All Sections (Devore)

SESSION 12 (LIVE IN-PERSON)

Probability

Topic 4. Probability

Problem Solving

SESSION 13 (LIVE IN-PERSON)

Probability

Topic 4. Probability

Problem Set 2

MODULE 3: RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

SESSION 14 (LIVE IN-PERSON)

Review Modules 1 and 2

SESSION 15 (LIVE IN-PERSON)

MID-TERM EXAM

The mid-term exam covers the first two modules of the course:

Module 1. Descriptive Statistics

Module 2. Probability

Please note that this date is tentative. Depending on the class pace the midterm exam could take place either in session 14 or in session 16.

SESSIONS 16 - 18 (LIVE IN-PERSON)

Topic 5. Discrete Random Variable. Discrete Probability Distributions

Definition of random variable. Probability distributions for discrete random variables. Properties of discrete random variables: Expected value, variance. Bernouilli distribution. Binomial distribution, Hypergeometric distributions, Negative Binomial Distribution. Poisson distribution, etc.

Readings:

Book Chapters: Chapter 4. Sections 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 (NCT)

Book Chapters: Chapter 4. Sections 4.1, 4.2, 4.3, 4.4 (MBS)

Book Chapters: Chapters 3&4. Sections 3.4, 4.2, 4.3, 4.5 (OIS)

Book Chapters: Chapter 3. All Sections (Devore)

SESSION 19 (LIVE IN-PERSON)

RStudio

Probability Distributions with RStudio

Discrete Probability Distributions.

Calculating Binomial probabilities. Calculating Poisson probabilities. Calculating Hypergeometric probabilities, etc.

SESSION 20 (LIVE IN-PERSON)

Random Variables & Probability Distributions

Topic 5. Discrete Random Variable. Discrete Probability Distributions

Problem Set 3

SESSIONS 21 - 22 (LIVE IN-PERSON)

Topic 6. Continuous Random Variables. Continuous Probability Distributions

Continuous random variables. Expectations for continuous random variables. The Uniform distribution. The Normal distribution. Descriptive Methods for Assessing Normality. Normal distribution Approximation for Binomial distribution. Chi-Square distribution. t distribution. F distribution. Exponential distribution, etc.

Readings:

Book Chapters: Chapters: Ch.5: All Sections, Ch.6: 6.4, Ch.7: 7.3, Ch.10: 10.4 (NCT)

Book Chapters: Chapter 4. Sections 4.5, 4.6, 4.7, 4.8 (MBS)

Book Chapters: Chapters 3&4. Sections 3.4, 3.5, 4.1, 4.3, 6.3.3, 7.1.3 (OIS)

Book Chapters: Chapter 4. Sections 4.1, 4.2, 4.3, 4.4 (Devore)

SESSION 23 (LIVE IN-PERSON)

RStudio

Probability Distributions with RStudio

Continuous Probability Distributions.

Calculating Normal probabilities. Calculating Uniform probabilities. Calculating Exponential probabilities, etc.

SESSION 24 (LIVE IN-PERSON)

Topic 6. Continuous Random Variable. Continuous Probability Distributions

Problem Set 4

SESSIONS 25 - 26 (LIVE IN-PERSON)

Topic 7. Joint Probability Distributions

Jointly Distributed Discrete Random Variables.

Jointly Distributed Continuous Random Variables.

Readings:

Book Chapters: Chapters: Ch.4: Section 4.7. Ch.5: Section 5.6 (NCT)

Book Chapters: Chapter 5. Sections 5.1, 5.2 (Devore)

SESSION 27 (LIVE IN-PERSON)

Topic 7. Joint Probability Distributions

Problem Set 5

SESSION 28 (LIVE IN-PERSON)

General Review

SESSION 29 (LIVE IN-PERSON)

Computer exam

The computer exam covers all the topics of the course

SESSION 30 (LIVE IN-PERSON)

Final Exam

The final exam covers all the topics of the course

EVALUATION CRITERIA

Your final grade in the course will be based on a combination of different items that are described in the following table:

criteria	percentage	Learning Objectives	Comments
Midterm exam	20 %		
Computer exam	20 %		
Final Exam	30 %		
Individual Work	20 %		
Class Participation	10 %		

RE-SIT / RE-TAKE POLICY

Class participation

Active participation: participation in class will be evaluated positively if students: (1) attain a threshold quantity of contributions that is sufficient for making a reliable assessment of comment quality. Additionally, (2) participation will be evaluated in quality terms. A high-quality comment reveals a depth of insight, rigorous use of case evidence, consistency of argument, and realism. A high-quality presentation of ideas must consider the relevance and timing of comments and the flow and content of the ensuing class discussion. It demands comments that are concise and clear, and that are conveyed with a spirit of involvement in the discussion at hand.

Problem Sets

Students will solve four-five problem sets throughout the course. These problem sets will be composed of theory and practice questions. These problem sets should be solved by hand and/or using RStudio. The aim of these problem sets is that students understand complex statistical concepts through the resolution of exercises. Answers will be submitted through BlackBoard using the format 'multiple-answer test'. Questions will be composed of several options of which approximately one third will be correct. The resolution of these problem sets will help students better prepare for the the midterm exam and the final exam. The problem set with the lowest score will be discarded.

Mid-term exam

The mid-term exam will take place in session 15. Please note that this date is tentative. Depending on the class pace the midterm exam could take place either in session 14 or in session 16. The format of this exam will be quite similar to that of the problem sets. The exam is closed book. Students must bring their own simple calculator (phones, tablets, laptops, and other electronic devices are not allowed). Students are also allowed to bring up one double-sided A4 SHEET paper with any formulae considered helpful. The use of Respondus is mandatory.

Computer exam

The computer exam consists in solving and discussing some questions using RStudio. This exam covers all the topics of the course and it is an open book exam.

Final Exam

The final exam will be similar to the problem sets, theory and practice questions to be submitted via BlackBoard using the format multiple-answer test. The final exam is scheduled in session 30, and will cover all the content of the course. The final exam will include material from the book, the PowerPoint slides, the problem sets, and class notes. It is highly recommendable to delve deeply into the topics using the books and the resources used during the course. The final exam is closed book. Students must bring their own simple calculator (phones, tablets, laptops, and other electronic devices are not allowed). Students are also allowed to bring up two double-sided A4 SHEETS paper with any formulae considered helpful. The use of Respondus is mandatory.

Passing grade

Two requirements must be met to pass the course:

1. The weighted average of the computer exam and the theory final exam has to be at least 3.5, i.e, (0.4*Grade Computer exam)+(0.6*Grade Final exam)>=3.5. If the weighted average does not reach the threshold value of 3.5, you will fail the course.

2. Your weighted average (computed using all the items in the table above) is at least 5

Retake exam

Those students who failed the subject in the first regular period will have to do a retake in July (except those not complying with attendance rules who are banned from this possibility).

Dates and location of the July retakes will be posted in advance and will not be changed. Please take this into consideration when planning your summer.

The maximum grade that a student may obtain in the 2nd exam session is 8 out of 10.

As per University 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 70% 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:

1. 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).

2. 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.)

3. 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.

4. 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

- Paul Newbold, William L. Carlson, Betty M. Thorne. (2022). *Statistics for Business and Economics.* 10th. Pearson. ISBN 9781292436845 (Digital)

Recommended

- Devore, Jay L.. (2016). *Probability and Statistics for Engineering and the Sciences.* 9th. Cengage Learning. ISBN 9781305251809 (Digital)

- David M Diez, Mine Çetinkaya-Rundel, Christopher D Barr. (2019). *OpenIntro Statistics.* 4th. OpenIntro, Inc. ISBN 1943450072 (Digital)

This book is available in an online version only and can be acquired via https://www.openintro.org/book/os/

- Mcclave, J.T, Benson, P.G., & Sincich, T.. (2018). *Statistics for Business and Economics.* 13th. Pearson Education Limited. ISBN 9781292227085 (Digital)

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.