FUNDAMENTALS OF PROBABILITY AND STATISTICS

Dual Degree in Business Administration & Data and Business Analytics BBADBA SEP-2024 FPS-NBDA.1.M.A

Area Applied Mathematics

Number of sessions: 30

Academic year: 24-25

Degree course: FIRST

Number of credits: 6.0

Semester: 1º

Category: BASIC

Language: English

Professor: **ANDREA CREMASCHI**

E-mail: acremaschi@faculty.ie.edu

Andrea Cremaschi, PhD

A Bayesian Statistician by training, Dr. Cremaschi is specialised in Biostatistics and Bayesian nonparametric methodology.

After obtaining his Ph.D. in Statistics at the University of Kent (UK), he joined the Biostatistics Department at the University of Oslo (Norway) as a Post-doctoral fellow. There, he developed a strong interest for applications in the field of biomedical studies and Biostatistics, while maintaining a passion for methodological Bayesian Statistics. He continued his career in the research institute A*STAR, in Singapore, where he had the chance to apply novel statistical methodologies to clinical studies relative to diabetes and mental health. He has published in international scientific journals and has supervised thesis for students at different level of their journey.

Office Hours

Office hours will be on request. Please contact at:

acremaschi@faculty.ie.edu

SUBJECT DESCRIPTION

Uncertainty and data are everywhere. Understanding data and making sense of the numbers we encounter every day, whether you're looking at survey results, sports scores, or business reports, is an extremely valuable asset to have in your skillset.

This course provides an introduction to the fundamental concepts of probability theory and statistics. Students will learn the basic principles underlying the analysis of data and the interpretation of uncertainty. Topics covered include descriptive statistics, probability distributions, hypothesis testing. Emphasis will be placed on both theoretical understanding and practical application through hands-on exercises and real-world examples.

The contents of this course will form 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, among others.

COURSE OUTLINE

MODULE I – Intro and Descriptive Statistics

Important definitions and practical examples. Data types. Measures of centrality. Measures of dispersion. Skewness and kurtosis. Graphical representation of data.

MODULE II - Probability

Definition of probability. Random experiments, events and sets. Basic probability rules. Conditional probability and independence. Counting techniques: permutations and combinations.

MODULE III - Probability Distributions

Discrete probability distributions. Continuous probability distributions. Properties of probability distributions. Expected value, variance and moments. Relationship between known distributions. Approximation of probability distributions.

MODULE IV - Sampling Distributions and Estimation

Population, sample and sampling. Properties of the samples. Sampling distributions of sample statistics.

MODULE V - Hypothesis Testing

Null and alternative hypotheses. Type I and Type II errors. Z-tests and t-tests. P-values and significance levels. Test for mean, variance and proportions.

LEARNING OBJECTIVES

At the end of the course, the students will have the necessary tools to analyse data in a variety of problems and business applications, and will be able to:

- Understand the basic principles of probability theory and its applications.
- Learn how to describe and summarise data of different type using appropriate statistical measures.
- Link probability distributions to patterns of randomness encountered in real-life applications (e.g., business and social activities).
- Use the most common distribution tables.
- Understand the meaning and implications of fundamental theoretical results of Statistics and Probability, such as the Central Limit Theorem and the Law of Large numbers.
- Translate practical questions into statistical term and apply the correct inferential procedure to provide a quantitative answer (relative to the types of problems encountered in class).
- Gain proficiency in conducting hypothesis tests and interpreting their results.

- Apply probability and statistical concepts to real-world problems and decision-making scenarios.

TEACHING METHODOLOGY

The course consists of 30 sessions. The sessions will be provided in two formats: in-person or asynchronous, meaning that some sessions will be held in class while others will be completed by each student individually outside class hours.

In-person: These sessions will follow the traditional classroom format. You will come to the classroom at the designated time for lecture.

· Theory sessions: The main resources in these classes will be projections of slides, providing files and electronic information and the use of the Internet as a support tool, work, and communication. The students must supplement these lectures with recommended readings by the professor, which will be made available in advance.

· Practical: During these sessions, the students will work on exercises and case studies tightly connected with the theoretical concepts explained during the theoretical sessions. These are tutored classes where the professor will guide the students through the steps needed to solve the posed questions. These problems will be available in advance. Most of the sessions and projects will require computations, so students should be familiarised with the software R (freely available). Bringing your laptop is mandatory to all sessions. If some of the exercises are not covered due to time constraints, the students are strongly encouraged to solve them in their own study time by using the theory lectures and the recommended bibliography, as they constitute examinable material.

Students are expected to participate actively in class, expressing their difficulties and engaging in discussions when appropriate. The professor will always help the students by clarifying the problem sets solutions; assuming the students have previously tried to work on their own. Otherwise, this will be counterproductive.

Learning Activity	**Weighting**	**Estimated time a student should dedicate to prepare for and participate in**
Lectures	30.0 %	45.0 hours
Discussions	3.3 %	5.0 hours
Exercises in class, Asynchronous sessions, Field Work	20.0 %	30.0 hours
Group work	20.0 %	30.0 hours
Individual studying	26.7 %	40.0 hours
TOTAL	100.0 %	150.0 hours

AI POLICY

In today's world, generative artificial intelligence (GenAI) is changing how we work, study and, in general, how we get things done. However, in the context of this course, the use of GenAI is not permitted, unless it is otherwise stated by the instructor. The use of GenAI tools would jeopardize the students' ability to acquire fundamental knowledge or skills of this course.

If a student is found to have used AI-generated content for any form of assessment, it will be considered academic misconduct, and the student might fail the respective assignment or the course.

PROGRAM

SESSION 1 (LIVE IN-PERSON)

Introduction: what is uncertainty and why do we care about it? Important definitions and practical examples. Organisation of the course and material.

SESSION 2 (LIVE IN-PERSON)

MODULE I – Descriptive Statistics

Data types. Measures of central tendency: mean, median, mode. Measures of dispersion: range, variance, standard deviation. Skewness and kurtosis. Quantiles, Covariance, Correlation, Contingency Tables. Graphical representation of data: Histograms, Scatter Plots, Boxplots. Chebyshev's Rule and the Empirical rule. The sample z-score.

SESSION 3 (LIVE IN-PERSON)

Brief intro to R/R Studio.

Data cleaning and preparation: loading data, tables, counting, missing values, plots, etc.

SESSION 4 (LIVE IN-PERSON)

MODULE I – Descriptive Statistics

Data types. Measures of central tendency: mean, median, mode. Measures of dispersion: range, variance, standard deviation. Skewness and kurtosis. Quantiles, Covariance, Correlation, Contingency Tables. Graphical representation of data: Histograms, Scatter Plots, Boxplots. Chebyshev's Rule and the Empirical rule. The sample z-score. Other concepts.

SESSION 5 (LIVE IN-PERSON)

MODULE II – Probability

Basic definitions. Random experiment, events and sets. Unions and intersections. The Additive Rule and the Multiplicative Rule. Conditional probability: Bayes Theorem. Random experiments. Independence. Joint probability. Counting techniques: permutations and combinations. Other concepts.

SESSION 6 (LIVE IN-PERSON)

Descriptive Statistics with R. Producing and interpreting graphs.

SESSION 7 (LIVE IN-PERSON)

MODULE III - Probability Distributions

Random variables. Expected value, variance, moments. Discrete random variables: probability functions and properties. Discrete Uniform, Bernoulli, Binomial, Poisson, Geometric and Hypergeometric probability distributions.

SESSION 8 (LIVE IN-PERSON)

MODULE III - Probability Distributions

Random variables. Expected value, variance, moments. Discrete random variables: probability functions and properties. Discrete Uniform, Bernoulli, Binomial, Poisson, Geometric and Hypergeometric probability distributions.

SESSION 9 (LIVE IN-PERSON)

Probability exercises.

SESSION 10 (LIVE IN-PERSON)

MODULE III - Probability Distributions

Continuous random variables: definition and properties. Expected value, variance, moments.

Uniform, Normal, and Exponential distributions. Descriptive Methods for Assessing Normality. Relationship between known distributions. Approximation of probability distributions.

SESSION 11 (LIVE IN-PERSON)

Quiz 1 (MODULE I and MODULE II).

SESSION 12 (LIVE IN-PERSON)

Probability Distributions with R.

Project.

SESSION 13 (LIVE IN-PERSON)

MODULE III - Probability Distributions

Continuous random variables: definition and properties. Expected value, variance, moments.

Uniform, Normal, and Exponential distributions. Descriptive Methods for Assessing Normality. Relationship between known distributions. Approximation of probability distributions.

SESSION 14 (LIVE IN-PERSON)

MODULE III - Probability Distributions

Continuous random variables: definition and properties. Expected value, variance, moments.

Uniform, Normal, and Exponential distributions. Descriptive Methods for Assessing Normality. Relationship between known distributions. Approximation of probability distributions. Extra: t-Student distribution, Gamma/Chi squared distributions.

SESSION 15 (LIVE IN-PERSON)

Problem Set Discrete Random Variable.

****

Exercises on Discrete Random Variables.

SESSION 16 (LIVE IN-PERSON)

MODULE IV - Sampling Distributions and Estimation

Population and sample. Random sampling and random samples. Properties: unbiasedness and minimum variance. The Law of Large Numbers. The Central Limit Theorem. Sampling Distribution of the Sample Proportion.

SESSION 17 (LIVE IN-PERSON)

MODULE IV - Sampling Distributions and Estimation

Confidence Interval for a Population Mean: Normal Statistic. t-Statistic. Confidence Interval for a Population Proportion. Determining the sample size. Confidence Interval for a Population Variance.

SESSION 18 (LIVE IN-PERSON)

Problem Set Continuous Random Variables.

****

Exercises on Continuous Random Variables.

SESSION 19 (LIVE IN-PERSON)

MODULE V - Hypothesis Testing

Concepts of Hypothesis Testing. Test of the Mean of a Normal Distribution: population variance known and population variance unknown. The t-distribution. The concept of p-value. Test of the population proportion (large samples). The Power of a test. Sample sizes. Test of Hypothesis about a population variance.

SESSION 20 (LIVE IN-PERSON)

MODULE V - Hypothesis Testing

Concepts of Hypothesis Testing. Test of the Mean of a Normal Distribution: population variance known and population variance unknown. The t-distribution. The concept of p-value. Test of the population proportion (large samples). The Power of a test. Sample sizes. Test of Hypothesis about a population variance.

SESSION 21 (LIVE IN-PERSON)

Problem Set Sampling Distributions.

****

Exercises on Sampling Distributions.

SESSION 22 (LIVE IN-PERSON)

Quiz 2 (MODULE III & MODULE IV).

SESSION 23 (LIVE IN-PERSON)

MODULE V - Hypothesis Testing

Concepts of Hypothesis Testing. Test of the Mean of a Normal Distribution: population variance known and population variance unknown. The t-distribution. The concept of p-value. Test of the population proportion (large samples). The Power of a test. Sample sizes. Test of Hypothesis about a population variance. Tests for proportions.

SESSION 24 (LIVE IN-PERSON)

Inferential Statistics with R.

SESSION 25 (LIVE IN-PERSON)

Problem set of Confidence Intervals.

SESSION 26 (LIVE IN-PERSON)

MODULE V - Hypothesis Testing

Concepts of Hypothesis Testing. Test of the Mean of a Normal Distribution: population variance known and population variance unknown. The t-distribution. The concept of p-value. Test of the population proportion (large samples). The Power of a test. Sample sizes. Test of Hypothesis about a population variance. Tests for proportions.

SESSION 27 (LIVE IN-PERSON)

Project presentations.

SESSION 28 (LIVE IN-PERSON)

Problem set Hypotheses Tests for a Single Population.

SESSION 29 (LIVE IN-PERSON)

General Review. Q&A.

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.

Class participation

Three main criteria will be used in reaching a judgment about your class participation:

1. Attendance: Attendance to class is compulsory. (1) Students must comply with the 80% attendance rule. (2) Punctuality will be taken into consideration when grading this assistance item and the teacher reserves the right to (not) allow attendance to class to those students not being on time. Finally, (3) general attitude and behaviour in class will be also considered. Students affecting the class environment in a negative way will lose points in the assistance grade.
2. 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.Exercises: Students may solve exercises remotely and check their comprehension of the different topics. Students will receive feedback about their submitted answers in an easy way. Grades will be based on the performance of these exercises and the time devoted to their resolution. It is highly recommended that you solve these exercises by hand and using R.

Quizzes

Quizzes will be announced in advance, and they will be computer based. There will be two quizzes.

Group project

The elements for effective teamwork can help you to achieve success in all areas of life. It is crucial that you learn how to collaborate with other students even when you have not chosen the students whom to work with. Each group will be composed of four students and must prepare a project due at the end of the course (more details about the final deadline will be periodically provided during the course.) The group project will consist in the identification of a real-world problem, taken from social sciences or any other field of interest, the collection of relevant data, the statistical analysis of the data, the development of a statistical model, and the final interpretation of the obtained results. Submissions will be delivered using Turnitin, following the appropriate link provided on campus online. No work will be accepted if submitted otherwise. At the beginning of the course, the professor will upload a specific document with project instructions where a step-by-step description of what the students need to do.

Final Exam

The final exam will be a written exam to be taken in session 30 and will cover all the content of the course. The final exam will include material from the PowerPoint slides and the problem sets. It is highly recommendable to delve deeply into the topics using the books and the resources used during the course.

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-sided A4 SHEET paper with any formulae considered helpful. Students must bring the Stat-tables, available on campus online. 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.

criteria	**percentage**	**Learning Objectives**	**Comments**
Final Exam	35 %		
Group Presentation	25 %		
Class Participation	10 %		
Intermediate tests	30 %		

RE-SIT / RE-TAKE POLICY

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 4 chances to pass any given course distributed in two consecutive academic years (regular period and July period). It is mandatory to attend 100% of the classes. Students who do not comply with at least 80% attendance will lose their 1st and 2nd chance and go directly to the 3rd one (they will need to enrol again in this course the next academic year). Grading for retakes will be subject to the following rules:

1. 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).

2. 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.

3. The maximum grade that a student may obtain in the 2nd exam session is 8 out of 10. Those students in the 3rd call will be required to attend 50% of the classes. If due to schedule overlap, a different option will be discussed with the professor to pass the subject.

BIBLIOGRAPHY

Compulsory

- Casella, G. and Berger, R.L.. (2002). *Statistical Inference.* 2nd. Pacific Grove (California) Duxbury Press. ISBN 9780534243128 (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.