FUNDAMENTALS OF DATA ANALYSIS

Grado en Computación e Inteligencia Artificial / Bachelor in Computer Science and Artificial Intelligence BCSAI SEP-2025
FDA-N-CSAI.1.M.A

Area Mathematics
Number of sessions: 30
Academic year: 25-26
Degree course: FIRST
Number of credits: 6.0
Semester: 2º
Category: BASIC
Language: English

Professor: **SIMÓN ISAZA**

E-mail: sisaza@faculty.ie.edu

Simón Isaza has a PhD. In mathematical research by Complutense University of Madrid. He completed these studies founded by an FPI scholarship and graduated “cum laude”. His thesis “CW Decompositions of Algebraic Curves and Milnor Fibers of Non-Isolated Quasi-Ordinary Singularities” treated about subjects in Topology and Theory of Singularities. He has worked in teaching and research at the National University of Colombia, Complutense University, and currently at IE University. He also has experience as financial risk consultant at KPMG. He is passionate about Technology, Economics, and knowledge in general.

Office Hours

Office hours will be on request. Please contact at:

sisaza@faculty.ie.edu

SUBJECT DESCRIPTION

The use of probability models and statistical methods for analyzing data has become common practice in virtually all scientific disciplines. Statistical concepts and methods are not only useful but indeed often indispensable in understanding the world around us. They provide ways of gaining new insights into the behavior of many phenomena that you will encounter in your chosen field of specialization in science.

Data analysis teaches us how to make intelligent judgments and informed decisions in the presence of uncertainty and variation. Without uncertainty or variation, there would be little need for statistical methods. If every component of a particular type had exactly the same lifetime, if human behavior were to lead to the same decision making, then a single observation would reveal all required information. Data analysis helps us to deal with uncertainty. Data analysis offers not only methods for analyzing the results of experiments once they have been carried out but also suggestions for how experiments can be performed in an efficient manner to mitigate the effects of variation and have a better chance of producing correct conclusions.

In this course, students will learn how to make inferences using statistics, or functions of observed data. They will perform confidence intervals and hypothesis testing, using one and two samples, analysis of variance, with one factor or more factors; and inferential statistics using categorical data.

This course will be the basis for further subjects as Algorithms and Data Structures, Probability for Computer Science; AI: Statistical Learning and Prediction, among others.

The use of probability models and statistical methods for analyzing data has become common practice in virtually all scientific disciplines. Statistical concepts and methods are not only useful but indeed often indispensable in understanding the world around us. They provide ways of gaining new insights into the behavior of many phenomena that you will encounter in your chosen field of specialization in science.

Data analysis teaches us how to make intelligent judgments and informed decisions in the presence of uncertainty and variation. Without uncertainty or variation, there would be little need for statistical methods. If every component of a particular type had exactly the same lifetime, if human behavior were to lead to the same decision making, then a single observation would reveal all required information. Data analysis helps us to deal with uncertainty. Data analysis offers not only methods for analyzing the results of experiments once they have been carried out but also suggestions for how experiments can be performed in an efficient manner to mitigate the effects of variation and have a better chance of producing correct conclusions.

In this course, students will learn how to make inferences using statistics, or functions of observed data. They will perform confidence intervals and hypothesis testing, using one and two samples, analysis of variance, with one factor or more factors; and inferential statistics using categorical data.

This course will be the basis for further subjects as Algorithms and Data Structures, Probability for Computer Science; AI: Statistical Learning and Prediction, 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 disciplines: computer sciences, engineering, physics, etc. At the end of the course; students should acquire:

- Proficiency to analyze and synthetize the main information content in a set of univariate and multivariate data.
- Proficiency to compute probabilitiesUnderstanding key concepts related to hypothesis testing;
- Proficiency to use random variables to model real phenomena

- Proficiency to perform inferences in one and two populations
- Proficiency in testing hypotheses about populations
- Proficiency in designing experiments and run analysis of variance
- Proficiency in dealing with categorical data

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

- The ability to think analytically
- The use of statistical software and programming language, namely Python
- The ability to think critically.

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 disciplines: computer sciences, engineering, physics, etc. At the end of the course; students should acquire:

- Proficiency to analyze and synthetize the main information content in a set of univariate and multivariate data.
- Proficiency to compute probabilitiesUnderstanding key concepts related to hypothesis testing;
- Proficiency to use random variables to model real phenomena
- Proficiency to perform inferences in one and two populations
- Proficiency in testing hypotheses about populations
- Proficiency in designing experiments and run analysis of variance
- Proficiency in dealing with categorical data

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

- The ability to think analytically
- The use of statistical software and programming language, namely Python
- 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	30.0 %	45.0 hours
Discussions	3.3 %	5.0 hours
Exercises in class, Asynchronous sessions, Field Work	30.0 %	45.0 hours
Group work	6.7 %	10.0 hours
Individual studying	30.0 %	45.0 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)

SESSION 1 [THEORY]: PRESENTATION AND TOPIC 1

Topic 1: Sampling distributions

Statistics and their distributions. Random samples. Deriving a sampling distribution. Simulation experiments. The distribution of the sample mean. The Central Limit Theorem. The distribution of a linear combination.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 2 (LIVE IN-PERSON)

SESSION 2 [THEORY]: TOPIC 1

Topic 1: Sampling distributions

Statistics and their distributions. Random samples. Deriving a sampling distribution. Simulation experiments. The distribution of the sample mean. The Central Limit Theorem. The distribution of a linear combination.

SYNCHRONOUS Live classes / lectures

Readings:

SYNCHRONOUS Live classes / lectures

Suggested materials Session 2
Book Chapters : Chapter 5. Sections 5.3, 5.4, 5.5 (DEV)

SESSION 3 (LIVE IN-PERSON)

SESSION 3 [THEORY]: TOPIC 2

Topic 2: Point Estimation

Given a parameter of interest, such as population mean or population proportion, the objective of point estimation is to use a sample to compute a number that represents in some sense an educated guess for the true value of the parameter. The resulting number is called a point estimate. This topic describes and illustrates two important methods for obtaining point estimates.

General concepts of point estimation. Methods of point estimation. Method of moments. Method of maximum likelihood estimation.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 4 (LIVE IN-PERSON)

SESSION 4 [THEORY]: TOPIC 2

Topic 2: Point Estimation

Given a parameter of interest, such as population mean or population proportion, the objective of point estimation is to use a sample to compute a number that represents in some sense an educated guess for the true value of the parameter. The resulting number is called a point estimate. This topic describes and illustrates two important methods for obtaining point estimates.

General concepts of point estimation. Methods of point estimation. Method of moments. Method of maximum likelihood estimation.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 5 (LIVE IN-PERSON)

SESSION 5 [THEORY]: TOPIC 2

Topic 2: Point Estimation

Given a parameter of interest, such as population mean or population proportion, the objective of point estimation is to use a sample to compute a number that represents in some sense an educated guess for the true value of the parameter. The resulting number is called a point estimate. This topic describes and illustrates two important methods for obtaining point estimates.

General concepts of point estimation. Methods of point estimation. Method of moments. Method of maximum likelihood estimation.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 6 (LIVE IN-PERSON)

SESSION 6 [PRACTICE]: Sampling distributions and Point Estimation

Problem Solving

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 7 (LIVE IN-PERSON)

SESSION 7 [THEORY]: TOPIC 3

Topic 3: Statistical Intervals Based on a Single Sample

A point estimate, because it is a single number, by itself provides no information about the precision and reliability of estimation. Al alternative to reporting a single sensible value for the parameter being estimated is to calculate and report an entire interval of plausible values -a confidence interval. Information about the precision of an interval estimate is conveyed by the width of the interval.

Basic properties of confidence intervals. Large-Sample confidence intervals for a population mean and proportion. Intervals based on a Normal population distribution. Confidence intervals for the variance and standard deviation of a Normal population.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 8 (LIVE IN-PERSON)

SESSION 8 [THEORY]: TOPIC 3

Topic 3: Statistical Intervals Based on a Single Sample

A point estimate, because it is a single number, by itself provides no information about the precision and reliability of estimation. Al alternative to reporting a single sensible value for the parameter being estimated is to calculate and report an entire interval of plausible values -a confidence interval. Information about the precision of an interval estimate is conveyed by the width of the interval.

Basic properties of confidence intervals. Large-Sample confidence intervals for a population mean and proportion. Intervals based on a Normal population distribution. Confidence intervals for the variance and standard deviation of a Normal population.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 9 (LIVE IN-PERSON)

SESSION 9 [PRACTICE]: Statistical Intervals Based on a Single Sample

Problem Solving

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 10 (LIVE IN-PERSON)

SESSION 10 [PRACTICE]: One sample inference with Python

SYNCHRONOUS Live classes / lectures

SESSION 11 (LIVE IN-PERSON)

SESSION 11 [PRACTICE]

Problem Set 1

SESSION 12 (LIVE IN-PERSON)

SESSION 12 [THEORY]: TOPIC 4

Topic 4: Tests of Hypotheses based on a single sample.

Frequently the objective of an investigation is not to estimate a parameter but to decide which of two contradictory claims about the parameter is correct. Methods for accomplishing this comprise the part of statistical inference called hypothesis testing.

Hypotheses and test procedures. z tests for hypotheses about a population mean. The one-sample t test. Tests concerning a population proportion. Tests concerning a population variance. Further aspects of hypothesis testing.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 13 (LIVE IN-PERSON)

SESSION 13 [THEORY]: TOPIC 4

Topic 4: Tests of Hypotheses based on a single sample.

Frequently the objective of an investigation is not to estimate a parameter but to decide which of two contradictory claims about the parameter is correct. Methods for accomplishing this comprise the part of statistical inference called hypothesis testing.

Hypotheses and test procedures. z tests for hypotheses about a population mean. The one-sample t test. Tests concerning a population proportion. Tests concerning a population variance. Further aspects of hypothesis testing.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 14 (LIVE IN-PERSON)

SESSION 14 [PRACTICE]: Tests of Hypotheses based on a single sample

Problem Solving

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 15 (LIVE IN-PERSON)

SESSION 15 [PRACTICE]: Tests of Hypotheses based on a single sample with Python

Hypothesis Testing with Python: One sample

SYNCHRONOUS Live classes / lectures

SESSION 16 (LIVE IN-PERSON)

SESSION 16 [THEORY]: TOPIC 5

Topic 5: Inferences based on two samples

This topic presents confidence intervals and tests for making inferences about a difference between two population parameters.

z tests and confidence intervals for a difference between two population means. The two-sample t test and confidence interval. Analysis of paired data. Inferences concerning a difference between population proportions. Inferences concerning two population variances.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 17 (LIVE IN-PERSON)

SESSION 17 [THEORY]: TOPIC 5

Topic 5: Inferences based on two samples

This topic presents confidence intervals and tests for making inferences about a difference between two population parameters.

z tests and confidence intervals for a difference between two population means. The two-sample t test and confidence interval. Analysis of paired data. Inferences concerning a difference between population proportions. Inferences concerning two population variances.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 18 (LIVE IN-PERSON)

SESSION 18 [PRACTICE]

Problem Set 2

SESSION 19 (LIVE IN-PERSON)

SESSION 19 [THEORY]: TOPIC 5

Topic 5: Inferences based on two samples

This topic presents confidence intervals and tests for making inferences about a difference between two population parameters.

z tests and confidence intervals for a difference between two population means. The two-sample t test and confidence interval. Analysis of paired data. Inferences concerning a difference between population proportions. Inferences concerning two population variances.

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 20 (LIVE IN-PERSON)

SESSION 20: MIDTERM EXAM (Exam via BlackBoard Ultra)

Module 1: Topics 1, 2, 3 & 4

Please note that this date is tentative. Depending on the class pace the midterm exam could take place either in session 18 or in session 19

SYNCHRONOUS Live classes / lectures

SESSION 21 (LIVE IN-PERSON)

SESSION 21 [PRACTICE]: Inference based on two samples

Problem Solving

SYNCHRONOUS Live classes / lectures

Readings:

SESSION 22 (LIVE IN-PERSON)

SESSION 22 [PRACTICE]: Inference based on two samples with Python

SYNCHRONOUS Live classes / lectures

SESSION 23 (LIVE IN-PERSON)

SESSION 23 [THEORY]: TOPIC 6

Topic 6: Analysis of Variance. Single Factor

ANOVA refers broadly to a collection of experimental situations and statistical procedures for the analysis of quantitative responses from experimental units. The simplest ANOVA problem is referred to as single factor or one-way ANOVA. It involves the analysis either of data sampled from more than two numerical populations (distributions) or of data from experiments in which more than two treatments have been used.

Single-Factor ANOVA. Multiple Comparisons in ANOVA. More on Single-Factor ANOVA.

SYNCHRONOUS Live classes / lectures

SESSION 24 (LIVE IN-PERSON)

SESSION 24 [PRACTICE]: ANOVA. Single factor with Python

SYNCHRONOUS Live classes / lectures

SESSION 25 (LIVE IN-PERSON)

SESSION 25 [PRACTICE]

Problem Set 3

SESSION 26 (LIVE IN-PERSON)

SESSION 26 [THEORY]: TOPIC 7

Topic 7: Multifactor Analysis of Variance

In many experimental situations, there are two or more factors that are of simultaneous interest. This topic extends the methods of topic 5 to investigate such multifactor situations.

Two-Factor ANOVA. The Fixed Effects Model. Randomized Block Experiments. Models with Random and Mixed Effects.

SYNCHRONOUS Live classes / lectures

SESSION 27 (LIVE IN-PERSON)

SESSION 27 [PRACTICE]: ANOVA Multifactor with Python

SYNCHRONOUS Live classes / lectures

SESSION 28 (LIVE IN-PERSON)

SESSION 28 [PRACTICE]

Problem Set 4

SESSION 29 (LIVE IN-PERSON)

SESSION 29 [PRACTICE]: Computer exam with Python

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

SYNCHRONOUS Live classes / lectures

SESSION 30 (LIVE IN-PERSON)

SESSION 30 [PRACTICE]: Final exam

The final exam covers all the content of the course, from the first session to the last session. The final exam will be similar to the problem sets, theory and practice questions.

EVALUATION CRITERIA

Ordinary evaluation

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

criteria	percentage	Learning Objectives	Comments
Class Participation	10 %		
Problem sets	20 %		
Midterm exam	20 %		

Computer exam	20 %		
Final Exam	30 %		

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 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 Python. 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, the computer exam, and the final exam.

Mid-term exam

The mid-term exam will take place in session 20. Please note that this date is tentative. Depending on the class pace the midterm exam could take place either in session 18 or in session 19. The format of this exam will be quite similar to that of the problem sets. 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 handwritten with any formulae considered helpful. The midterm exam is a closed book exam. Communication by any means is strictly forbidden. Students must connect to the Internet using the IE network, otherwise the grade of the quiz will be 0.

Computer exam

The computer exam consists in solving and discussing some questions using Python. This exam covers all the topics of the course and it is an open book exam. This exam will be taken individually and it is intended to evaluate students' understanding of the material and the use of Python. Communication by any means is strictly forbidden. Students must connect to the Internet using the IE network, otherwise the grade of the quiz will be 0

Final Exam

It will be taken in session 35 and will cover all the content of the course, from the first session to the last session. The final exam will be similar to the problem sets, theory and practice questions. 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 all the resources used during the course. The final exam is closed book. Students must bring their own simple calculator (neither programmable nor graphical calculators are allowed. Phones, tablets, laptops, and other electronic devices are not allowed). Students are also allowed to bring up two-sided A4 SHEET paper handwritten with any formulae considered helpful. No questions are allowed during the exams. Communication by any means is strictly forbidden. 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.

Resit/Retake policy

Each student has four chances to pass any given course distributed over two consecutive academic years: ordinary call exams and extraordinary call exams (re-sits) in June/July.

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

Evaluation criteria:

- Students failing the course in the ordinary call (during the semester) will have to re-sit the exam in June / July (except those not complying with the attendance rule, who will not have that opportunity and must directly re-enroll in the course on the next Academic Year).
- The extraordinary call exams in June / July (re-sits) require your physical presence at the campus you are enrolled in (Segovia or Madrid). There is no possibility to change the date, location or format of any exam, under any circumstances. Dates and location of the June / July re-sit exams will be posted in advance. Please take this into consideration when planning your summer.
- The June / July re-sit exam will consist of a comprehensive exam. Your final grade for the course will depend on the performance in this exam only; continuous evaluation over the semester will not be taken into consideration. Students will have to achieve the minimum passing grade of 5 and can obtain a maximum grade of 8.0 (out of 10.0) – i.e., “notable” in the in the re-sit exam.
- Retakers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will be needed to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as retakers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained in the retake exam (3rd call) is 10.0.

After ordinary and extraordinary call exams are graded by the professor, you will have a possibility to attend a review session for that exam and course grade. Please be available to attend the session in order to clarify any concerns you might have regarding your exam. Your professor will inform you about the time and place of the review session. Any grade appeals require that the student attended the review session prior to appealing.

- Students failing more than 18 ECTS credits in the academic year 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 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 extra cost. As you know, students have a total of four allowed calls to pass a given subject or course, in order to remain in the program.

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.