FUNDAMENTALS OF DATA ANALYSIS

ASIGNATURAS CONVERGENTES
Professor: BORJA MESA SANCHEZ
E-mail: bmesa@faculty.ie.edu

Academic year: 18-19
Degree course: FIRST
Semester: 2º
Category: COMPULSORY
Number of credits: 6.0
Language: English

PREREQUISITES
Fundamentals of probability and statistics, Data insights and visualization.

SUBJECT DESCRIPTION
Statistics is a discipline grounded on mathematics. We use mathematical tools to construct formal models to help us understand how the world is. In Statistics we use mathematical tools to transform raw data into useful information. Statistics helps us to summarize and process data. The aim is to obtain accurate estimates by using inferential statistics, as a way of enhancing the decision-making process by reducing uncertainty. Statistics are used to reduce uncertainty in many different environments: financial markets, insurance industry, biomedicine, consumer behavior, presidential elections, gambling industry, physics, etc.

At the end of the course, students will have learnt how to make inference using statistics, or functions of observed data. Students will learn how to estimate uncertain parameters (point and interval estimation), how to test statistical hypothesis, how to compare two populations and finally, the main concepts of the linear regression models.

This course will be the basis for further subjects as Probability and Statistics for Data Management and Analysis and others.

OBJECTIVES AND SKILLS
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 business, social and behavioral applications. At the end of the course students should be able to:

- Perform statistical inference in one and two populations
- Understanding of key concepts related with hypothesis testing
- Design experiments and the analysis of variance
- Testing categorical probabilities.
- Perform linear regression models.
- Carry out the previous analysis by using statistical software

Additionally, the course will focus on the acquisition or reinforcement of generic skills:
- The ability to think analytically.
- The use of statistical software
- The ability to think critically

**OBJECTIVES AND SKILLS**

**METHODOLOGY**

The course consists of 30 sessions. There will be a balance between theoretical and practical sessions.

Lectures: The main resources in these classes will be projections of slides, providing files and electronic information and the use of Internet as a support tool, work and communication. The student must supplement these lectures with recommended readings by the professor. All of them will be available to students.

Practical: They focus on exercises, case studies and practical cases specified well in advance and that are temporarily connected with the theoretical concepts taught. Students must prepare and solve the problem sets assigned prior to the beginning of the practice class by using the theory lectures and the recommended bibliography. Students are expected to participate actively in class, expressing their difficulties and proposing solutions. The professor will clarify problem sets solutions, but this will be productive only if students have previously tried to work on their own. Some problems and projects will require computations, so students should be familiarized with software R.

**METHODOLOGY**

<table>
<thead>
<tr>
<th>Teaching methodology</th>
<th>Weighting</th>
<th>Estimated time a student should dedicate to prepare for and participate in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures</td>
<td>20.0 %</td>
<td>30 hours</td>
</tr>
<tr>
<td>Discussions</td>
<td>6.67 %</td>
<td>10 hours</td>
</tr>
<tr>
<td>Exercises</td>
<td>30.0 %</td>
<td>45 hours</td>
</tr>
<tr>
<td>Group work</td>
<td>30.0 %</td>
<td>45 hours</td>
</tr>
<tr>
<td>Other individual studying</td>
<td>13.33 %</td>
<td>20 hours</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100.0 %</td>
<td>150 hours</td>
</tr>
</tbody>
</table>
PROGRAM

SESSIONS 1 - 3 (FACE TO FACE)
UNIT 1: INFERENCES BASED ON A SINGLE SAMPLE (Sessions 1 to 8)

PRESENTATION & TOPIC 1
Topic 1: Inferences based on a single sample: Estimation with confidence intervals
Presentation of the programme, aims and guidelines of the course. Point estimation of parameters. Confidence interval for a population mean (large samples, normal distribution). Confidence interval for a population mean (small samples, t distribution). Large sample confidence interval for a population proportion. Determining the sample size.
B.C.: Chapter 6 (Sections 6.1 to 6.5)
Case Study 1

SESSION 4 (LABORATORY)
Exercises

SESSIONS 5 - 7
TOPIC 2
Topic 2: Inferences based on a single sample: Test of hypotheses
The elements of a test of hypothesis. Formulating null and alternative hypothesis. Observed significance levels: p-values. Test of hypothesis about a population mean: normal (z) statistic. Test of hypothesis about a population mean: Student’s t-statistic. Large-sample test of hypothesis about a population proportion. Type II error.
B.C.: Chapter 7 (Sections 7.1 to 7.6, 7.8)
Case Study 2

SESSION 8 (LABORATORY)
Exercises

SESSIONS 9 - 11
UNIT 2: INFERENCES BASED ON TWO SAMPLES (Sessions 9 to 12)

TOPIC 3
Topic 3: Inferences based on two samples: Confidence intervals and Tests of hypotheses
B.C.: Chapter 8 (Sections 8.1 to 8.5)
Case Study 3

SESSION 12 (LABORATORY)

25th October 2018
SESSIONS 13 - 16
UNIT 3: DESIGN OF EXPERIMENTS AND ANALYSIS OF VARIANCE (Sessions 13 to 17)

TOPIC 4
Topic 4: Design of experiments and analysis of variance
Elements of a designed experiment. The completely randomized design: single factor. Multiple comparisons of means. The randomized block design. Factorial experiments: two factors.
*B.C.: Chapter 9 (Sections 9.1 to 9.5)*
Case Study 4

SESSION 17 (LABORATORY)
Exercises

SESSION 18 (LABORATORY)
Review of units 1,2 and 3

SESSION 19 (LABORATORY)
Mid-term exam

SESSIONS 20 - 21
UNIT 4: CATEGORICAL DATA ANALYSIS (Sessions 20 to 22)

TOPIC 5
Topic 5: Categorical Data Analysis
*B.C.: Chapter 10 (Sections 10.1, 10.2, 10.3)*
Case Study 5

SESSION 22 (LABORATORY)
Exercises

SESSIONS 23 - 25 (FACE TO FACE)

TOPIC 6
Topic 6: Simple linear regression
Probabilistic models. The least squares approach. Assumptions of the model. Inference about the slope. Coefficient of determination. Estimation and prediction. A complete example
*B.C.: Chapter 11 (Sections 11.1 to 11.7)*
Case Study 6

25th October 2018
SESSION 26 (LABORATORY)
Exercises

SESSIONS 27 - 31

TOPIC 7
Topic 7: Multiple regression
Multiple regression models. First-order models with quantitative independent variables. Model building in multiple regression. Multiple regression diagnostics.

B.C.: Chapter 12
Case Study 7

SESSION 32 (LABORATORY)
Exercises

SESSION 33 (FORUM)
Presentation of group projects

SESSION 34 (FACE TO FACE)
General review

SESSION 35 (LABORATORY)
Final exam
BIBLIOGRAPHY

All the required readings are from the compulsory textbook “Statistics for Business and Economics”, McClave J.T, Benson P.G. & Sincich T. Pearson Prentice Hall, 13th edition, 2018 (ISBN / ISSN: 978-1-29-222713-9). Reading a section means reading the text AND the examples. You are required to obtain a copy and read the indicated sections in advance, i.e., before each lecture.

The following is a recommended book, “Open Intro Statistics”, Diez, David; Barr Christopher & Cetinkaya-Rundel, Mine, 3rd edition. This book can be used to review some of the topics presented in class, to find extra exercises, etc. The textbook is offered under a Creative Commons license at: https://www.openintro.org/

Visit openintro.org for a free PDF, to download the textbook’s source files, or for more information about the license.

EVALUATION CRITERIA

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Percentage</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Exam</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>Intermediate Tests</td>
<td>20%</td>
<td>Mid-term</td>
</tr>
<tr>
<td>Group Presentation</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>Intermediate Tests</td>
<td>10%</td>
<td>Quizzes</td>
</tr>
<tr>
<td>Class Participation</td>
<td>10%</td>
<td></td>
</tr>
</tbody>
</table>

A. CLASS PARTICIPATION (10%)

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

1- Assistance: Assistance to class is compulsory. (1) Students must comply with the 70% attendance rule. Otherwise they will lose their 1st and 2nd chance, and go directly to the 3rd one (they will need to enroll again in this course next academic year). (2) Punctuality will be taken into consideration when grading this assistance item and the teacher reserves the right to 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. A high quality comment reveals 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.

B. GROUP REPORT (20%)

Each group should be composed of 4 to 5 students and must prepare a group report due at the end of 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.

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. 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.
C. QUIZZES (10%)
At the beginning of some sessions, you will be given a short quiz based on required readings and exercises for the session.

D. MID-TERM EXAM (20%)
The mid-term exam will take place around session 19 and will cover Units 1, 2, and 3 of the content of the course. Notice that the date of the midterm could change and need to be considered with flexibility. The precise date will be communicated to students two weeks ahead of time.

You must bring your own simple calculator (phones, tablets, laptops and other electronic devices are not allowed). You are also allowed to bring up one-sided A4 SHEET paper with any formulae that you think could be helpful. NO QUESTIONS ARE ALLOWED DURING THE EXAM. THE CHEAT-SHEET ALONG WITH ANY SCRAP PAPER WILL BE COLLECTED AND STAPLED TO YOUR EXAM.

E. FINAL EXAM (40%)
For the final exam, you must bring your own simple calculator (phones, tablets, laptops and other electronic devices are not allowed). You are also allowed to bring up two-sided A4 SHEET paper with any formulae that you think could be helpful. NO QUESTIONS ARE ALLOWED DURING THE EXAMS. THE CHEAT-SHEET ALONG WITH ANY SCRAP PAPER WILL BE COLLECTED AND STAPLED TO YOUR EXAMS.

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.

RETAKE POLICY
Each student has 4 chances to pass any given course distributed over two consecutive academic years: ordinary call exams and extraordinary call exams (re-sits) in June/July.

Any student whose weighted final grade is below 5 will be required to sit for the retake exam to pass the course. Students who do not comply with the 70% 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 will be subject to the following rules:
- 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 the maximum grade will be capped at 8.0 (out of 10.0) – i.e., “notable” in the in the re-sit exam.
- 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 in the retake exam (3rd call) is 10.0 (out of 10.0).
- The non-July retakes (this happens in the ordinary period: students in their third attempt) will entail a midterm and a final exam. The weights are as follows: midterm 40%, final 60%. In order to pass, a minimum of 3.5 in the final exam is required.
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.

- 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 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 4 allowed calls to pass a given subject or course, in order to remain in the program.

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

**PROFESSOR BIO**

Professor: **BORJA MESA SANCHEZ**  
E-mail: bmesa@faculty.ie.edu

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 Masters 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

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

**BIBLIOGRAPHY**

**CODE OF CONDUCT IN CLASS**

1. **Be on time**: Students arriving more than 5 minutes late will be marked as “Absent”.

25th October 2018
Only students that notify in advance in writing that they will be late for a specific session may be granted an exception (at the discretion of the professor).

2. **If applicable, bring your name card and strictly follow the seating chart.** It helps faculty members and fellow students learn your names.

3. **Do not leave the room during the lecture:** Students are not allowed to leave the room during lectures. If a student leaves the room during lectures, he/she will not be allowed to re-enter and, therefore, will be marked as “Absent”.

   Only students that notify that they have a special reason to leave the session early will be granted an exception (at the discretion of the professor).

4. **Do not engage in side conversation.** As a sign of respect toward the person presenting the lecture (the teacher as well as fellow students), side conversations are not allowed. If you have a question, raise your hand and ask it. If you do not want to ask it during the lecture, feel free to approach your teacher after class.

   If a student is disrupting the flow of the lecture, he/she will be asked to leave the classroom and, consequently, will be marked as “Absent”.

5. **Use your laptop for course-related purposes only.** The use of laptops during lectures must be authorized by the professor. The use of Social Media or accessing any type of content not related to the lecture is penalized. The student will be asked to leave the room and, consequently, will be marked as “Absent”.

6. **No cellular phones:** IE University implements a “Phone-free Classroom” policy and, therefore, the use of phones, tablets, etc. is forbidden inside the classroom. Failing to abide by this rule entails expulsion from the room and will be counted as one absence.

7. **Escalation policy: 1/3/5.** Items 4, 5, and 6 above entail expulsion from the classroom and the consequent marking of the student as “Absent.” IE University implements an “escalation policy”: The first time a student is asked to leave the room for disciplinary reasons (as per items 4, 5, and 6 above), the student will incur one absence, the second time it will count as three absences, and from the third time onward, any expulsion from the classroom due to disciplinary issues will entail 5 absences.