PROBABILITY & STATISTICS FOR DATA MANAGEMENT AND ANALYSIS

BACHELOR IN DATA AND BUSINESS ANALYTICS
Professor: RAFIF SROUR DAHER
E-mail: rsrour@faculty.ie.edu

Academic year: 19-20
Degree course: SECOND
   Semester: 1º
Category: BASIC
Number of credits: 6.0
Language: English

PREREQUISITES
SUBJECT DESCRIPTION

Nowadays, all companies are striving to become more profitable, to reach customers quicker, and to offer higher-quality products and services. In addition, businesses want to reach these objectives with less fewer and at lower costs. It seems like an impossible task, but it is not. An essential requirement in this process is effective knowledge creation and management. There is no lack of information, but there is a dearth of knowledge. As Tom Peters said in his book Thriving on Chaos, “we are drowning in information and starved for knowledge”.

In recent year, there has been an exponential increase in the amount of information available for decision making. Big data, as it is commonly known is currently being collected and stored in data warehouses, ready to be mined for actionable insights. Some of that data can be analyzed and understood with simple statistics, but much of it requires more complex, multivariate statistical techniques to convert into knowledge. Even more, some data requires to use new approaches such as the Bayesian statistics to transform it into usable data.

In this context, this course will cover basic multivariate analysis, both dependent and independent techniques such as factor analysis, multiple discriminant analysis, logistic regression, canonical correlation, conjoint and cluster analysis. An introduction to Bayesian theory is also covered.
OBJECTIVES AND SKILLS
The objective of this course is to provide students with a number of advanced statistical techniques needed for data analysis and management. Each dataset is unique and necessitates a special treatment. Understanding the nature of the dataset, its underlying structure and distribution, the objectives for which it has been collected are essential in determining the appropriate multivariate analysis tool.

At the end of the course; students should be able to:

- Compare and contrast parametric and non-parametric tests;
- Perform a number of non-parametric tests and identify multiple applications where their use is adequate;
- Explain what multivariate analysis is and when its applications are appropriate;
- Understand the six-step approach to multivariate model building;
- Determine which multivariate technique is suitable and for what purpose;
- Develop a solid understanding of Bayesian statistics; from theoretical foundations to computation and implementation.
METHODOLOGY

The course is both conceptual and practical; the theory explained in class goes hand-in-hand with application exercises using SPSS as the statistical software.

Conceptual sessions are delivered using Power Point presentations and focus on discussing the basic concepts underlying statistical theories along with various examples (test your understanding, worked examples, and challenging problems).

During the practical sessions; SPSS will be used to solve specific questions presented in the context of case-studies. Bringing your laptop is mandatory to all sessions, although its use (or not) will be decided by the professor.

Prior to all sessions, you should read assigned textbook sections. Reading the textbook in advance will allow you to get the most out of each lecture. When reading the textbook sections prior to each lecture, you must look at the examples but you do not need to solve them.

On weekly basis, a brief quiz covering previously taught material will be given. These quizzes are meant to test your overall understanding of the material and will help the professor assess the overall performance and evolution of the class.

Problem sets will be uploaded to campus online. I strongly recommend that you do the exercises given as homework during the course and not leave them for a date close to the exam.

Students are encouraged to work in groups when solving homework problems. But, as evaluation will be strictly individual, it is highly recommended that each student try, at the same time, to solve problems by himself/herself.
<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>10.0 %</td>
<td>15 hours</td>
</tr>
<tr>
<td>Exercises</td>
<td>20.0 %</td>
<td>30 hours</td>
</tr>
<tr>
<td>Group work</td>
<td>20.0 %</td>
<td>30 hours</td>
</tr>
<tr>
<td>Other individual studying</td>
<td>30.0 %</td>
<td>45 hours</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100.0 %</td>
<td>150 hours</td>
</tr>
</tbody>
</table>
PROGRAM

CONTENT
The theoretical content of this course consists of three parts and each part is divided into several units. The first part covers nonparametric tests namely Mann Whitney U test, Sign test and Wilcoxon Signed Rank Test. Kruskall Wallis test is also covered. The second part discusses multivariate analysis techniques while the third part is dedicated to the Bayesian theory.

All the required readings are from the compulsory textbooks “Multivariate data Analysis”, Hair, Black, Babin & Anderson, Pearson New International Edition and “The Bayesian Choice: From Decision-Theoratic Foundations to Computational Implementation”, 2nd ed. Springer. Reading a section means reading the text AND doing the examples.

Note: The following description of the material covered is tentative. An attempt will be made to cover all listed topics. However; the pace in the classes will depend on the group performance.

SESSIONS 1 - 2

Topics: Introduction and presentation of the course syllabus and objectives. Statistical significance versus statistical power: Review.

Required Reading
- Textbook: Multivariate Data Analysis: Chapter 1. pp 8 - 11

SESSION 3

Topics: Parametric versus non-parametric tests. Introduction to non-parametric testing.

Required Reading

SESSIONS 4 - 8

Topics: Mann Whitney U Test. Tests with matched samples: The sign Test and Wilcoxon Signed rank Test. Tests with more than two independent samples: Kruscal Wallis Test.

Required Reading

SESSION 9

REVIEW UNITS 1 & 2

SESSION 10

PARTIAL EXAM 1

SESSIONS 11 - 12

Required Reading

SESSION 13
Topics: Data assessment and evaluation. Missing values and outliers.
Required Reading

SESSIONS 14 - 17
Required Reading

SESSION 18
REVIEW UNITS 3, 4, & 5

SESSION 19
PARTIAL EXAM 2

SESSIONS 20 - 24
Required Reading

SESSIONS 25 - 27
Required Reading
- Textbook: The Bayesian Choice: Chapter 1.

SESSION 28
Student Series

SESSION 29
FINAL REVIEW

SESSION 30
BIBLIOGRAPHY

RECOMMENDED

Title: Multivariate Data Analysis
Authors: Joseph, F. Hair Jr, William C. Black, Barry J. Babin & Ralph E. Anderson.
Medium: PRINT ELECTRONIC

Title: The Bayesian Choice
Authors: Robert P. Christian.
Publisher / Edition / Year: Springer / 2nd edition / 2007
ISBN 10: 0387715983
ISBN 13: 9780387715988
Medium: PRINT ELECTRONIC

A free online version of Multivariate Data Analysis can be downloaded from: https://is.muni.cz/el/1423/podzim2017/PSY028/um/_Hair__Multivariate_data_analysis_7th_revised.pdf

Additional documents, including lecture slides, problem sets, extra readings, etc… will be posted on Campus Online throughout the semester.

To download the Non-parametric Tests handout, visit the following site: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric_print.html
EVALUATION CRITERIA

Your final grade in the course will be based on both individual and group work of different characteristics that will be weighted in the following way:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Percentage</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Participation</td>
<td>10 %</td>
<td></td>
</tr>
<tr>
<td>Workgroups</td>
<td>20 %</td>
<td></td>
</tr>
<tr>
<td>Intermediate Tests</td>
<td>30 %</td>
<td>15% EACH</td>
</tr>
<tr>
<td>Final Exam</td>
<td>25 %</td>
<td></td>
</tr>
<tr>
<td>Quizzes</td>
<td>15 %</td>
<td></td>
</tr>
</tbody>
</table>

A. Class participation and discussion

Class participation will be evaluated based on the following criteria:

- Quality (not quantity) of your participation in class discussion: The most important dimension of participation concerns what it is that you are saying. A high quality comment reveals depth of insight, rigorous use of case evidence, consistency of argument, and realism. Frequency refers to the attainment of a threshold quantity of contributions that is sufficient for making a reliable assessment of comment quality. The logic is simple: if contributions are too few, one cannot reliably assess the quality of your remarks. However, once threshold quantity has been achieved, simply increasing the number of times you talk does not automatically improve your evaluation. Beyond the threshold, it is the quality of your comments that must improve. In particular, one must be especially careful that in claiming more than a fair share of “airtime”, quality is not sacrificed for quantity. Finally, your attempts at participation should not be such that the instructor has to “go looking for you”. You should be attempting to get into the debate on a regular basis.

You might want to avoid being classified as one of the following types of students:

- Repeaters, i.e., students that, consciously or unconsciously, make comments that are really just repeats/rephrasing of what has already been said (by other students, or you). This wastes time and adds nothing to learning.
- Ramblers, i.e., students that take a lot of time to say simple things or they may tell long personal/professional stories, or they roam into topics that are not relevant, or simply make low-quality comments just to participate. They waste valuable time and prevent other students from being able to participate.
- Students that have been distracted (by Facebook, etc.) or who have stopped paying attention and then, later on, when they realized they have missed a term or concept, they ask you about it.

B. Group project_Student Series

The group project is an integral part of this course. Each group (randomly composed of 4 -5 students) will be asked to choose an advanced topic in Statistics, to write a paper on that topic and to prepare a presentation. These presentations will be delivered as part of a new initiative entitled: Students Series. This initiative is launched by BDBA to promote students communication and presentations skills in conveying complex concepts.
C. Quizzes
Throughout the semester and on weekly basis (unless otherwise specified by your instructor), you will be given a short announced online-quiz based on previously covered material. These quizzes will help you assess your overall understanding of the subject being studied and identify any caveat in your learning. **NO MAKE UP FOR QUIZZES WILL BE PERMITTED.** The dates of these quizzes will be set by the professor the first week of classes.

D. Exams
There will be two partial exams and one final exam. For these exams, 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 for each partial exam and 3 one-sided A4 sheets to the final exam. The sheets can only contain 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.

Notice that the date of the partial exams could change and need to be considered with flexibility. The precise date will be communicated to students two weeks ahead of time. The date of the final and the retake exam CANNOT BE CHANGED under any circumstances.

- The July retakes will consist of a comprehensive exam. The grade will depend only on the performance in this exam; continuous evaluation over the semester will not be taken into account. This exam will be designed bearing in mind that the passing grade is 5 and the maximum grade that can be attained is 8.
- 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.** Any changes to this policy will be communicated to students ahead of time.
- The maximum grade that a student may obtain in any type of retake will be **8 out of 10.**

**PROFESSOR BIO**
Professor: RAFIF SROUR DAHER
E-mail: rsrour@faculty.ie.edu

Fifteen years of experience in teaching, analytical and empirical research, consultancy, and data analysis. Demonstrated ability to work in international/multicultural environments (Lebanon, USA, Spain). Life-long learner; both academically and personally. Five times winner of Best Professor award and strong advocate of using coaching and mentoring to help students improve their academic performance and overall college experience.

EXPERIENCE
Dec 2018-Present Co-Director of Duals in Politics, Law, Economic and Data & Business Analytics at IE University, HST (Madrid, Spain)
Dec 2018-Present Co-Director of Duals in Business Administration & Business Analytics at IE University, HST (Madrid, Spain)
Dec 2017-Present Director of Bachelor in Data & Business Analytics at IE University, HST (Madrid, Spain)
Sep 2014-Present Professor-Instituto de Empresa (Madrid, Spain)
Jan 2010-Jan2011 Market Consultant-Perspectiv Project Management (Madrid, Spain)
Mar 2007-Nov 2007 Training Consultant–Chamber of Commerce, Industry and Agriculture in collaboration with the Order of Engineers (Lebanon)
Dec 2006-Mar 2007 Technology Implementation Consultant-American University of Beirut (Lebanon)
Sep 2005-Dec 2005 Technical Advisor-Rehabilitation of Lebanese Administration UNPD Lebanon)
Aug 2004-Aug 2005 Visiting Professor-West Virginia University (West Virginia, USA)
Jan 2001-April 2004 Academic Researcher–American University of Beirut (Lebanon)
Feb 1999-Dec 2000 Departmental Researcher–American University of Beirut (Lebanon)
Sep 1998-Aug 2000 Supervising Engineer–United Nation Development Program (Lebanon)
Sep 1998-Feb 1999 Graduate Research Assistant-American University of Beirut (Lebanon)

EDUCATION
2000 – 2004 PhD in Environmental Soil Chemistry West Virginia University, (USA)
1998 – 2000 MS in Soil Science American University of Beirut (Lebanon)
1994 – 1998 BE in Agricultural Engineer American University of Beirut (Lebanon)
1994 – 1998 BS in Agriculture American University of Beirut (Lebanon)

OTHER INFORMATION

19th June 2019
WHEN QUESTIONS ARISE OUT OF CLASS:

Email:
If you have a question(s) that was not answered in class, you are welcome to ask your question(s) via email. I can be reached at: rsrour@faculty.ie.edu Although I will make every effort to respond to your question(s) as quickly and thoroughly as possible, please recognize that I may not be available when you send an email. Thus, please allow me up to 48 hours to respond before sending a follow-up email.

Office Hours:
If your question cannot be properly answered via email and/or you would prefer to meet in person, please make an appointment to meet with me on the university campus during my scheduled office hours. Office hours will be determined at the beginning of the semester and posted on Campus Online.

OTHER INFORMATION:

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 70% attendance will lose their 1st and 2nd chance, and go directly to the 3rd one (they will need to enroll 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 in order to pass the subject.

Attendance:
Attendance at all scheduled classes is mandatory and essential for success in the course. If you miss class for any reason, you are responsible for getting notes from classmates. If you have questions about any assignment please send me an email. Under most circumstances, students who miss a class in which a presentation, mid-term, or final exam is held will not be granted an exception or given an opportunity to do a make-up assignment or exam. However, if illness or other circumstances prevent you from adhering to the assignment/presentation due dates stated in this syllabus, an exception may be granted at the discretion of the professor. In all cases, the student must provide official documentation (e.g., from a medical doctor, counsellor) to the professor within 24 hours of the missed due date.

Students with Special Needs:
To request academic accommodations due to a disability, please contact Rafif Srour via email at: Rafif.Srour@ie.edu.

Student Privacy Statement:
At times, students may disclose personal information through class discussions. It is expected that all members of the class will respect the privacy of their classmates. This means that the information disclosed in the class will not be repeated or discussed with other students outside of the course.

Decisions about Grades:

19th June 2019
Decisions about grades are made very carefully, and are final at the end of the course. If you have questions regarding a certain grade or you would like to receive personal feedback, you must request a meeting with me to discuss grades on specific assignments before the last class of the course. Any disputes regarding grades must be resolved before the final exam. "Extra credit" or makeup assignments will only be allowed under extenuating circumstances at the sole discretion of the course professor.

**ACADEMIC INTEGRITY**

Unless you are specifically instructed to work with other students in a group, all of your assignments, papers, projects, presentations, and any work I assign must reflect your own work and thinking.

What is academic integrity? When you do the right thing even though no one is watching. The core values of integrity, both academic and otherwise include: honesty, fairness, respect, responsibility, and trust. Academic Integrity requires that all students within Instituto de Empresa (IE) act in accordance with these values in the conduct of their academic work, and that they follow the rules and regulations concerning the accepted conduct, practices and procedures of academic research and writing. Academic Integrity violations are defined as Cheating, Plagiarism or other violations of academic ethics.

Cheating and plagiarism are very serious offenses governed by the IE student code of conduct. Any student found cheating or plagiarizing on any assignment or component of this course will at a minimum receive a “0” on the affected assignment. Moreover, the student will also be referred to the University Judicial System for further action. Additional penalties could include a note on your transcript, failing the class, or expulsion from the university.

It is important to note that, while the list below is comprehensive, it should not be considered exhaustive.

**Cheating includes:**

a. An act or attempt to give, receive, share, or utilize unauthorized information or unauthorized assistance at any time for assignments, papers, projects, presentations, tests or examinations. Students are permitted to mentor and/or assist other students with assignments by providing insight and/or advice. However, students must not allow other students to copy their work, nor will students be permitted to copy the work of other students. Students must acknowledge when they have received assistance from others.

b. Failure to follow rules on assignments, papers, projects, presentations, tests or examinations as provided by the course professor and/or as stipulated by IE.

c. Unauthorized co-operation or collaboration.

d. Tampering with official documents, including electronic records.

e. The impersonation of a student on presentations, exercises, tests or an examination. This includes logging onto any electronic course management tool or program (e.g. Black Board, etc.) using someone else’s login and password.

**Plagiarism includes:**

a. Using the work of others and attempting to present it as your own. For example, using phrases or passages from books, articles, newspapers, or the internet and not referencing them properly in your document. This includes using information from others without citing it, misrepresentation of cited work, and misuse of quotation marks.

b. Submitting an assignment or paper that is highly similar to what someone else has written (i.e., minimal changes in wording, or where the sentences are similar, but in a different order).

c. You don’t have to commit “word for word” copying to plagiarize – you can also plagiarize if you turn in something that is “thought for thought” the same as someone else.

**Other violations of academic ethics include:**
a. Not acknowledging that your work or any part thereof has been submitted for credit elsewhere.
b. Misleading or false statements regarding work completed.

Knowingly aiding or abetting anyone in committing any form of an Academic Integrity violation.
CODE OF CONDUCT IN CLASS

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