Prof. Anna Sáez de Tejada Cuenca Prof. email: asaezdetejada@iese.edu Office: T-401 Phone: 506414	Assistant: Noemi Jerez Assistant email: njerez@iese.edu Office: T-500 Phone: 504465

STATISTICS

MRM 2023 1st Term

Introduction

This is a statistics course aimed at first-year doctoral students. This course has a hands-on (or learning-by-doing) approach, and it also relies heavily on data visualization. The course contains the fundamental concepts of probability, classical statistical inference, and a brief introduction to machine learning.

COVID-19 Policy

Everything in this course is subject to IESE's guidelines with regards to COVID-19, as well as the local, regional and national authorities' rules. Your safety, and your classmates', is the first priority. If you have any symptoms, please stay home and warn the PhD office, which will tell you how to proceed. In the event of a health-related circumstance, I will extend grace with regards to your attendance grade, your homework deadlines, etc.

Professor's Bio

Prof. Anna Sáez de Tejada Cuenca

Assistant Professor of Production, Technology and Operations Management Department

- Ph.D., in Operations Management, UCLA Anderson School of Management
- M.Sc., in Mathematical Engineering, Universitat Politècnica de Catalunya
- B.Sc., in Mathematics, Universitat Politècnica de Catalunya

Anna Sáez de Tejada Cuenca is an Assistant Professor of Production, Technology and Operations Management. She received her PhD in 2019 from the UCLA Anderson School of Management, and spent a year working as a Postdoctoral Research Fellow at Georgetown University's McDonough School of Business. Before starting her doctoral studies, Anna was a research assistant at IESE and a junior R&D engineer at Sabirmedical. During her PhD, she interned at Amazon in the Supply Chain Optimization Technologies team. Anna holds a BSc in Mathematics and a MSc in Mathematical Engineering from the School of Mathematics and Statistics at Universitat Politècnica de Catalunya.

Anna's research interests include social responsibility, sustainability, supply chain management, behavioral operations management, and empirical and experimental operations management. Recently she has been working on visibility and transparency in apparel supply chains, on behavioral biases of managers, on sustainable business models in fashion, on diffusion of diversity initiatives along supply chains, and on consumer attitudes and behavior with regards to environmental issues like recycling, renewable energies, or ecolabeled products. Her papers have been accepted in leading journals such as Management Science. Anna won the POMS College of Sustainable Operations Best Student Paper Award in 2018 and the POMS College of Behavior in Operations Management Junior Scholar Paper Competition in 2019.

Anna loves cats and string instruments, as well as riding her bicycle around the city.

Updated on September 9, 2021

Competences

- Deciding what the appropriate statistical tools are in different quantitative research contexts.
- Executing such tools using the statistical software R and its environment RStudio.
- Visualizing different types of data using the ggplot2 library in R.
- Communicating statistical results clearly and effectively.
- Designing and executing a simple quantitative research project.
- Understanding basic probability concepts and the probabilistic origin of all statistics tools.

Methodology

Each session will consist of a classic lecture, where I will be teaching you the concepts, tools, etc., and a hands-on part where everyone will use their laptop to start solving the homework assignment related to that session. During the hands-on part of the session, I will be walking around and assisting with any questions that may arise.

Expectations

I expect you to come to class well rested, hydrated, and fed, and ready to *be fully engaged* and willing to learn.

I expect you to *bring your laptop* to every session, *but keep it closed* until start the hands-on part of the class.

I expect you to *know the basics of R* before the course starts.

I expect you to submit your assignments *by their deadline* (I will grant extensions for force majeure causes).

I expect you to write and *turn in your own work* (code, math, and reports) for the homework assignments, even though you are allowed (and encouraged!) to discuss the problems with your classmates.

You can expect all the class materials made by me (homework assignments, notes, slides, etc.) to be posted on the Virtual Campus, usually right after their corresponding session.

You can expect me to announce any important information via the Virtual Campus's announcements tool, which also sends the announcement to your e-mail inboxes.

You can expect me to return your homework, with comments and grades, within a week of your submission.

You can expect me to be available to solve your questions related to the assignments or to the course content. There are no scheduled office yours, but you can e-mail me any time with questions or a request to meet up.

Recommendations

In this course you will be learning some statistics and visualization tools. So, imagine I was giving you physical tools in every session. Hopefully, you would be storing them very neatly in one of those toolboxes with compartments, so that the next time you needed a specific tool it would be easy for you to find it. With non-physical tools, such as formulas, pieces of code, etc., I recommend you do the same: store them neatly in your computer, because the toolkit you are building in this course will be useful for other times, such as other courses or your own research, and you will appreciate having organized your own tools so readily accessible.

I also recommend that, when you receive a graded assignment, you take some time to look at my comments and corrections, and submit a revised version of the homework. In the end, mistakes are the best teachers!

Evaluation

Attendance and engagement – 10%

This is not a discussion-based course like it would be in an MBA program. Therefore, I will not be evaluating "participation" as such. Having said this, I expect you to attend each session and be present (physically and mentally) and engaged, ask questions, help respond to your classmates' questions, etc.

Homework – 30%

This will be of two types (on the same week you may have homework of the two types):

R assignments: Hands-on practice of the methods you just learned. For each R homework, please *upload a small report (in PDF)* containing all your figures, tables, etc. as well as comments to the results you obtained. The report should not be particularly long, but it should display that you can interpret the graphics or the results of the tests you run. Please *upload your code (in a .R file)* in addition to the report.

Problem sets: Classic "pen and paper" exercises. You can either type them or write them on paper in legible handwriting. If you do the former, please *upload your document as a PDF*. If you do the latter, please *upload proof of having finished by the deadline* (i.e., a cell phone photo of your homework) and *turn in your hard copy* at the beginning of the following session.

Note that *you may revise your homework and I will regrade it*. In other words, based on my grade and comments, you can correct an assignment and submit it again (at the latest, on the evening before the final exam). If you do that, the grade that will count towards your final grade is that of the revised version.

Final project – 30%

The final project can be done individually or in pairs (evaluation standards will be set accordingly). You will come up with a small research question that can be answered using statistical methods. We will collect primary data (online) and you will analyze it and write a small paper with your findings. I will provide more detailed instructions in a separate document.

Final exam – 30%

It will have two parts: a short, closed-book test with open-text questions and problems (similar to the pen-and-paper homework), and a longer, open-book, hands-on part where you will run some statistical analyses on R.

Extra credit

Bad statistics show & tell: We will start some sessions by looking at an instance of bad (or intentionally misleading) usage of statistics or visualization (from a newspaper, TV, blog, social media, etc.). If you find a good example, please e-mail it to me before 20:00 on the day before class, and you will have the opportunity to present the example to your classmates. It can be in any language as long as you are able to translate it for us.

Help encoding the Final Project's surveys: If you know how to use Qualtrics, or will be using it for your research and want to learn how it works, you can help me encode your (and your classmates') surveys.

Other: I might offer extra credit for other types of extra work, such as asking challenging questions in class and then working independently to find and answer.

I will *multiply your final grade* by a number *between 1.0 and 1.1* based on how much extra credit you have earned.

Course Schedule

Please note that this may be subject to changes (I will warn you if that is the case).

Before Sep 16, 20:00	*Submit Homework 0*
Sep 17, 9:30–12:15	**Session 1: Descriptive statistics and data visualization** Required reading: Wheelan's chapters 1, 2, 3, 4
Before Sep 20, 20:00	*Submit Homework 1*
Sep 21, 14:30–17:15	**Session 2: Introduction to probability** Required reading: Wheelan's chapters 5, 5 ½
Before Sep 27, 20:00	*Submit Homework 2*
Sep 28, 14:30–17:15	**Session 3: Random variables and their most popular distributions** Required reading: Wheelan's chapter 6
Oct 1, 9:30–12:15	**Session 4: Estimation methods. Central limit theorem.** Required reading: Wheelan's chapters 7, 8
Before Oct 4, 20:00	*Submit Homework 3* *Submit Homework 4*
Oct 5, 14:30–17:15	**Session 5: Introduction to inference. Confidence intervals.** Required reading: Wheelan's chapters 9, 10
Before Oct 7, 20:00	*Submit Homework 5*
Oct 8, 9:30–12:15	**Session 6: Introduction to hypothesis testing**
Before Oct 14, 20:00	*Submit Homework 6*
Oct 15, 9:30–12:15	**Session 7: Non-parametric hypothesis testing**
Before Oct 18, 20:00	*Submit Homework 7*
Oct 19, 14:30–17:15	**Session 8: Introduction to linear regression** Required reading: Wheelan's chapters 11, 12, 13
Before Oct 21, 20:00	*Submit Homework 8*
Oct 22, 9:30–12:15	**Session 9: Introduction to machine learning** Required reading: Wheelan's conclusion chapter
Before Oct 25, 20:00	*Submit Homework 9*
Before Oct 27, 20:00	*Submit Final Project's proposal*
Before Oct 31, 20:00	I will return your proposal with comments and suggestions
Before Nov 2, 20:00	Pre-register Final Project's research design using aspredicted.org

Updated on September 9, 2021

	Submit proof of pre-registration
Nov 3 – Nov 7	I will encode your Final Projects' survey in Qualtrics (you can help me for extra credit!), launch it on Prolific, and upload the data once it has been collected.
Before Nov 22, 20:00	*Submit Final Project's report/paper* *(Optional) Submit revised homework assignments*
Nov 23, 14:30–17:15	**Final exam** Bring your laptop! You can also bring notes, books, etc. (but **not** another person to help you!)

Bibliography

Required books

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media.

Wheelan, C. (2013). *Naked statistics: Stripping the dread from the data*. WW Norton & Company.

Resources to learn R

Codeacademy. (n.d.). *Lear R.* https://www.codecademy.com/learn/learn-r/

Sievertsen, H. H. (2020). *Introduction to R.* https://hhsievertsen.shinyapps.io/r_introduction/

RStudio. (n.d.). *R cheatsheets*. https://www.rstudio.com/resources/cheatsheets/

Recommended books

That contain roughly this course's topics

Dalgaard, P. (2008). *Introductory Statistics with R*. Springer.

Freedman, D., Pisani, R., & Purves, R. (1998). *Statistics*. WW Norton & Company.

Navidi, W., & Monk, B. (2013). *Essential statistics*. McGraw-Hill Higher Education.

Hogg, R. V., McKean, J., & Craig, A. T. (2005). *Introduction to mathematical statistics*. Pearson Education.

To learn more about R and visualization using ggplot2

Teetor, P. (2011). *R cookbook: Proven recipes for data analysis, statistics, and graphics*. O'Reilly Media.

Chang, W. (2018). *R graphics cookbook: Practical recipes for visualizing data*. O'Reilly Media.

Healy, K. (2018). *Data visualization: a practical introduction*. Princeton University Press.

To learn about causal inference

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Pearl, J. (2009). *Causality*. Cambridge University Press.

To learn more about Bayesian statistics and machine learning

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer.

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.

To learn more about probability (some of these use very advanced real analysis)

Durrett, R. (2009). *Elementary probability for applications*. Cambridge University Press.

Ross, S. M. (2014). *Introduction to probability models*. Academic Press.

Hacking, I. (2016). *Logic of statistical inference*. Cambridge University Press.

Durrett, R. (2019). *Probability: theory and examples*. Cambridge University Press.

To learn more about the principles of data visualization

Tufte, E. R. (2001). *The visual display of quantitative information*. Graphics Press.

All the other books by Edward Tufte.

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

To learn how to write well

Dreyer, B. (2019). *Dreyer's English: An Utterly Correct Guide to Clarity and Style: The UK Edition*. Random House.

White, E. B., & Strunk, W. (1972). *The elements of style*. MacMillan.