Prof. Anna Sáez de Tejada Cuenca Prof. email: asaezdetejada@iese.edu Office: S-508 Phone: 506414	Assistant: Noemi Jerez Assistant email: njerez@iese.edu Office: T-500 Phone: 504465

STATISTICS

MRM 2024 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.

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.

Professor's Bio

Prof. Anna Sáez de Tejada Cuenca

Assistant Professor Operations, Information and Technology

• 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 Operations, Information and Technology. 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 sustainability, social responsibility, the circular economy, and behavioral operations management, from an empirical and experimental methodology. 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 July 26, 2022

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.

I expect you to *know 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 to be posted on the Virtual Campus.

You can expect me to announce any important information via the Virtual Campus's announcements tool.

You can expect me to post a solution key to your homework within a week of everyone having submitted them.

There are no scheduled office yours, but you can e-mail me any time with questions or a request to meet up.

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 – 10%

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

R assignments: Hands-on practice. Please *upload a small report (in PDF)* containing all your figures, tables, etc. as well as comments to the results you obtained. 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 (*upload your document as a PDF*) or handwrite them on paper (*upload proof of having finished by the deadline* — e.g. a cell phone photo of your homework — and *turn in your hard copy* at the beginning of the following session).

Final project – 40%

It 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. I will provide more detailed instructions in a separate document.

Final exam – 40%

Open-book, hands-on exam where you will solve some pen-and-paper questions and problems, and also 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 *add up to 10 points* to your final grade 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 15, 20:00	*Submit Homework 0*
Sep 16, 9:30–12:15	**Session 1: Introduction to probability.**
Before Sep 19, 20:00	*Submit Homework 1*
Sep 20, 14:30–17:15	**Session 2: Random variables.**
Before Sep 22, 20:00	*Submit Homework 2*
Sep 23, 14:30–17:15	**Session 3: Most popular probability distributions.**
Before Sep 26, 20:00	*Submit Homework 3*
Sep 27, 9:30–12:15	**Session 4: Estimation methods.**
Before Sep 29, 20:00	*Submit Homework 4*
Sep 30, 14:30–17:15	**Session 5: Introduction to inference. Central limit theorem. Confidence intervals.**
Before Oct 3, 20:00	*Submit Homework 5*
Oct 4, 9:30–12:15	**Session 6: Parametric hypothesis testing.**
Before Oct 6, 20:00	*Submit Homework 6*

Oct 7, 9:30–12:15	**Session 7: Non-parametric hypothesis testing.**
Before Oct 10, 20:00	*Submit Homework 7*
Oct 11, 14:30–17:15	**Session 8: Introduction to linear regression.**
Before Oct 13, 20:00	*Submit Homework 8*
Oct 14, 9:30–12:15	**Session 9: Introduction to machine learning**
Before Oct 14, 12:15	Discuss with me your idea for Final Project (nothing to submit, just talk)
Before Oct 20, 20:00	*Submit Final Project's proposal — first version*
Before Oct 24, 20:00	I will return your proposal with comments and suggestions
Before Oct 27, 20:00	*Submit Final Project's proposal — final version*
Before Oct 28, 20:00	Pre-register Final Project's research design using aspredicted.org *Submit proof of pre-registration (PDF generated by aspredicted.org)*
Oct 31 – Nov 4	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 21, 20:00	*Submit Final Project's report/paper*
Nov 22, 14:30–17:15	**Final exam** Bring your laptop! You can also bring notes, books, etc. (but **not** another person to help you!)

Bibliography

Resources to learn R

Codeacademy. *Learn R.* https://www.codecademy.com/learn/learn-r/

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

RStudio. *R cheatsheets*. https://www.rstudio.com/resources/cheatsheets/

Recommended books

That contain roughly this course's topics

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

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.

Fun book to learn more statistics (not a textbook)

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

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.

Software to help with symbolic calculus

Wolfram|Alpha. https://www.wolframalpha.com/