PROBABILITY AND STATISTICS

Master in International Development MID SEP-2024 S-1
Area International Relations
Number of sessions: 22
Term: Term 1
Category: regular
Language: English

Professor: **MICHAEL BECHER**

E-mail: mbecher@faculty.ie.edu

Michael Becher is Assistant Professor at IE University's School of Politics, Economics & Global Affairs. He obtained his PhD in Politics from Princeton University. Before joining IE University, he taught at the Toulouse School of Economics and the University of Konstanz. His research is situated in the fields of political economy and comparative politics. Broadly speaking, it uses quantitative methods to study the functioning of democracy and how to make it work better in the face of inequality, political polarization, and technological change. His work has been published in leading academic journals, such as American Journal of Political Science, American Political Science Review, or Journal of Politics and received multiple awards. It has also been featured in various international news outlets (e.g., Le Monde, The Conversation, or The Washington Post).

Office Hours
Office hours will be on request. Please contact at:

mbecher@faculty.ie.edu

SUBJECT DESCRIPTION

The course introduces the use of probability and statistics in the social sciences and international development in particular. The tremendous growth in data availability and technological progress means that scholars and decision makers in governments, international organizations, firms, or NGOs frequently turn to the quantitative analysis of data to make predictions and evaluate policies involving cause-and-effect relationships. Often, we want to know whether X causes Y, and by how much. Answering causal questions requires statistical concepts and tools. The course provides both.

LEARNING OBJECTIVES

The goals of this course are two-fold. A first objective of this course is to help students acquire the literacy for understanding research using quantitative data and statistical reasoning – to be good consumers of statistics and to identify the misuse of statistics and data in arguments. A second goal is to enable students to analyze data to address a particular research question, interpret the results, and effectively communicate the findings. After introducing the vocabulary to talk about causal research questions based on causal diagrams and potential outcomes, the course opens the statistical toolbox for the analysis of experimental and observational data. At a minimum, students should obtain the ability to calculate and interpret descriptive statistics for a given data set, understand and perform hypothesis tests about means and group differences as well as become able to interpret and conduct ordinary least squares regression for the analysis of experiments and observational data. Throughout the course, students will gain competence using R and RStudio, a statistical computation software which is freely available, open source, and widely used.

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: Readings, lectures, in-class data analysis activities (R labs), as well as individual and group work using R.

This is a "hands on" class – lectures are intended for summary and review of the assigned readings. In the computational lab sessions students will be instructed how to use R and RStudio, a popular computing software that is freely available and open source, to apply the techniques. The course is cumulative, and each class builds on what we have learned before. Whenever students miss a class, it is their responsibility learn what was covered in the session they missed.

Required Readings are indicated for each session in the program below. Completing the required reading prior to attending class is mandaroy and helps you learn the material quicker.

The course heavily draws on the following books. They can also serve as a companions in substantive courses, for theses, capstones, and for applied work beyond graduation:

- **Data Analysis**. Elena Llaudet and Kosuke Imai, Data Analysis For Social Science, Princeton University Press. Includes introduction to R and many examples.
- **Thinking Clearly**. Ethan Bueno de Mesquita and Anthony Fowler, Thinking Clearly with Data, Princeton University Press. More conceptual. Covers additional topics.
- **Impact Evaluation.** Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, M.J. (2016). Impact Evaluation in Practice. Second Edition. World Bank Group. Intuitive, many examples from field. Preprint available for free at https://openknowledge.worldbank.org/handle/10986/25030

Some sessions assign additional/other required readings from other sources. These can be applications in the form of an article or chapters from other books.

Optional Readings are not required. They may be covered during class and can cover an alternative exposition of the same material, additional examples, or more advanced treatment.

There are more books out there that people commonly consult to learn about these topics, including:

1. Angrist, Joshua David, and Jörn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press, 2009.

2. Huntington-Klein, N. (2021). The Effect: An Introduction to Research Design and Causality. CRC Press. Free online version available at: https://theeffectbook.net/

3. Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. New York: Basic Books.

4. Cunningham, S. (2021). Causal Inference: The Mixtape. New Haven: Yale University Press. Online version available at: https://mixtape.scunning.com/

Also, there are tons of video and audio resources about statistics and causal inference on the web. Here are some that I can recommend:

- Causal Inference Bootcamp (short introductory videos) https://mattmasten.github.io/bootcamp/
- Online Causal Inference Seminar https://sites.google.com/view/ocis/
- Causal Inference Podcast, https://casualinfer.libsyn.com/

In summary, the material we cover will come from a variety of sources, so lecture and lecture notes are an important source of information on which you will be examined. I encourage students to ask questions whenever they are in doubt – in class, in office hours, over email, etc. – and to participate in classroom activities. Lastly, the course website should be checked regularly for materials posted, announcements, and discussions.

Learning Activity	Weighting
Lectures	35.0 %
Exercises in class, Asynchronous sessions, Field Work	25.0 %
Group work	20.0 %
Individual studying	20.0 %
TOTAL	100.0 %

AI POLICY

Generative artificial intelligence (GenAI) tools may be used in this course for coding, with appropriate acknowledgement, as well as studying the course material. GenAI may not be used for in-class exams. If a student is found to have used AI-generated content inappropriately, it will be considered academic misconduct, and the student might fail the respective assignment or the course.

If used effectively, AI tools can significantly enhance your data analysis skills. However, they are no substitute for thinking hard about data and the question you want to answer using statistics. Using these and other tools effectively requires a sound understanding of foundational statistical concepts, data, and the coding environment you are working in. In this course and the following main course in Probability and Statistics, we are laying foundations.

Don't take ChatGPT's or any GenAI's output at face value. If you do not understand what the code is doing, do not assume it is correct. Assume it is wrong unless you either know the answer or can cross-check it with another source. You are responsible for any errors or omissions. You will be able to validate the outputs of GenAI for topics you understand.

AI is a tool, but one that you need to acknowledge using. Failure to do so is in violation of academic honesty policies. Acknowledging the use of AI will not impact your grade.

Suggested format to acknowledge the use of generative AI tools:

I acknowledge the use of [AI systems link] to [specify how you used generative AI]. The prompts used include [list of prompts]. The output of these prompts was used to [explain how you used the outputs in your work].

If you have chosen not to include any AI generated content in your assignment, the following disclosure is recommended:

No content generated by AI technologies has been used in this assignment.

PROGRAM

The program is orientative and likely to be somewhat in flux over the semester, as topics may take more or less than anticipated.

SESSION 1 (LIVE IN-PERSON)

Introduction

- Why evaluate?
- Causation: What is it and what it is good for?
- Potential outcomes and counterfactuals
- Fundamental problem of causal inference
- **Required reading:**

Book Chapters: Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, M.J.. (2016). Impact Evaluation in Practice - Chapters 1-2 (See Bibliography)

SESSIONS 2 - 3 (LIVE IN-PERSON)

Estimating causal effects with randomized experiments

- Theory and R lab
- Recap potential outcomes, individual and average causal effects
- What is an experiment?
- Randomized experiments and the difference-in-means estimator
- **Required reading**:
  - Data Analysis. Ch. 2.
- Also watch one of the following **videos** (between 30 minutes each):
  - Abhijit Banerjee, Prize lecture for Prize in Economic Sciences (YouTube)
  - Esther Duflo, Prize lecture for Prize in Economic Sciences on "Field Experiments and the Practice of Policy" (The Nobel Prize)
  - Michael Kremer, Prize lecture for Prize in Economic Sciences (The Nobel Prize)
- **Optional reading**:
  - James N. Druckman, Donald P. Green, James H. Kulinski and Arthur Lupia, (Eds.), Cambridge Handbook of Experimental Political Science - Chapter 2. Experiments: An Introduction to Core Concepts
  - Impact Evaluation. Ch. 3-4. (See link to free version below)

Book Chapters: Elena Llaudet and Kosuke Imai. (2023). Data Analysis For Social Science - Chapter 2 (See Bibliography)

Other / Complementary Documentation: Video. Abhijit Banerjee, Prize lecture for Prize in Economic Sciences (YouTube)

Other / Complementary Documentation: Video. Esther Duflo, Prize lecture for Prize in Economic Sciences on "Field Experiments and the Practice of Policy" (The Nobel Prize)

Other / Complementary Documentation: Video. Michael Kremer, Prize lecture for Prize in Economic Sciences (The Nobel Prize)

Book Chapters: James N. Druckman, Donald P. Green, James H. Kulinski and Arthur Lupia, (Eds.), Cambridge Handbook of Experimental Political Science - Chapter 2. Experiments: An Introduction to Core Concepts (CED)(Optional)

Book Chapters: Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, M.J.. (2016). Impact Evaluation in Practice - Chapters 3-4 (See Bibliography)(Optional)

SESSIONS 4 - 5 (LIVE IN-PERSON)

Inferring population characteristics using survey research

- Review of theory and R lab: Survey sampling vs. RCTs
- **Required reading:**
  - Data Analysis. Ch. 3.

Book Chapters: Elena Llaudet and Kosuke Imai. (2023). Data Analysis For Social Science - Chapter 3 (See Bibliography)

SESSIONS 6 - 7 (LIVE IN-PERSON)

Describing and predicting outcomes using linear regression

- Theory and R lab
- Line fitting and the linear regression model
- The least squares method
- Relationship to correlation coefficient
- Linear model, non-linear data
- **Required reading:**
  - Data Analysis, Ch. 4
- **Optional reading:**
  - Thinking Clearly, Ch. 4-5

Book Chapters: Elena Llaudet and Kosuke Imai. (2023). Data Analysis For Social Science - Chapter 4 (See Bibliography)

Book Chapters: Ethan Bueno de Mesquita and Anthony Fowler. (2021). Thinking Clearly with Data - Chapters 4-5 (See Bibliography)(Optional)

SESSIONS 8 - 9 (LIVE IN-PERSON)

Uncertainty and statistical inference I

- R ACTIVITY 1 DUE NIGHT BEFORE CLASS ON BLACKBOARD
- Population and sample
- Estimand and estimator
- Bias and consistency
- Standard error, 95% confidence interval, null hypothesis, statistical significance, p value

- **Required reading:**
    - Thinking Clearly, Ch. 6 (less technical)
    - Data Analysis, Ch. 6-7

Book Chapters: Ethan Bueno de Mesquita and Anthony Fowler. (2021). Thinking Clearly with Data - Chapter 6 (See Bibliography)
Book Chapters: Elena Llaudet and Kosuke Imai. (2023). Data Analysis For Social Science - Chapters 6-7 (See Bibliography)

SESSIONS 10 - 11 (LIVE IN-PERSON)

Uncertainty and statistical inference II

- Theory and R Lab, cont.
- Required reading: see above

SESSIONS 12 - 13 (LIVE IN-PERSON)

Estimating causal effects with observational data I

- Theory and R lab
- Why correlation does not imply causation
- Omitted variable bias and reverse causality
- How does regression control?
- Controlling and causation
- Controlling for confounders vs. mechanisms
- Required reading:
    - Data Analysis, Ch. 5
    - Thinking Clearly, Ch. 10.
- Optional reading:
    - Huntington-Klein, N., 2021. The Effect: An Introduction to Research Design and Causality. CRC Press. Ch. 6-8.

Book Chapters: Elena Llaudet and Kosuke Imai. (2023). Data Analysis For Social Science - Chapter 5 (See Bibliography)
Book Chapters: Ethan Bueno de Mesquita and Anthony Fowler. (2021). Thinking Clearly with Data - Chapter 10 (See Bibliography)
Book Chapters: Huntington-Klein, N., 2021. The Effect: An Introduction to Research Design and Causality. CRC Press - Chapters 6-8 (IE Library)(Optional)

SESSIONS 14 - 15 (LIVE IN-PERSON)

Group project and R Lab

SESSIONS 16 - 17 (LIVE IN-PERSON)

Estimating causal effects with observational data II

- Theory and R lab, cont.
- Required reading: see session I on topic
- Optional: Watch one of the videos on causal inference below:

Other / Complementary Documentation: Video. Interview with Donald Rubin (YouTube)

Other / Complementary Documentation: Video. Judea Pearl: The New Science of Cause and Effect (YouTube)

Other / Complementary Documentation: Video. Andrew Gelman: 100 Stories of Causal Inference (YouTube)

Other / Complementary Documentation: Video. Fireside chat with Susan Athey (YouTube)

Other / Complementary Documentation: Video. Guido Imbens (YouTube)

SESSION 18 (LIVE IN-PERSON)

MOCK EXAM

- Practice exam (not graded) and review

- R ACTIVITY 2 DUE NIGHT BEFORE CLASS ON BLACKBOARD

SESSIONS 19 - 20 (LIVE IN-PERSON)

Group project presentation

- Presentation of empirical group project in class

- Final replication material for final project is due one week after the presentation.

SESSION 21 (LIVE IN-PERSON)

Review session

- Review before the final exam

SESSION 22 (LIVE IN-PERSON)

EXAM

- In-class, closed book, individual

EVALUATION CRITERIA

Exam: The individual exam takes place in class during the last session; it is closed book. Please note that there will be a mock exam in advance.

R activities: Students will complete two R activities to build competence using R and to reinforce course concepts. While these assignments will be started during lab sessions, they should be completed outside of class as homework. Students are permitted to work in groups on the R activities at their own discretion, but students must individually submit their answers via Blackboard.

Group project: Working in groups, students identify a research question, and analyze appropriate data to try to answer it. The results are presented in a presentation in the last quarter of the course. The final replication material for the presentation is due one week after the presentation.

Class participation: Contributions in class (not mere attendance) such as questions, discussion; may also include quizzes on required readings and mini-presentations.

There are no extra credit assignmets.

criteria	percentage	Learning Objectives	Comments
Exam	40 %		In person, closed book

R activities	30 %		Homework
Group presentation	20 %		Empirical project
Class Participation	10 %		May include quizzes on required readings and mini-presentations

FAILING GRADE AND REASSESSMENT

When students receive a Fail in a course, they have the opportunity to present themselves for reassessment in order to earn the necessary credits toward graduation.

The reassessment of students should be scheduled between 5 and 10 working days after the review session takes place.

Grades for the reassessment are limited to a Low Pass and Fail.

Both, the initial Fail as well as the grade of the reassessment remain on the transcript. For the purpose of calculating the GPA however, only the grade of the reassessment is to be considered. Students receiving a failing grade in the reassessment of a course will not be able to continue in the program.

BIBLIOGRAPHY

Compulsory

 - Elena Llaudet and Kosuke Imai. (2023). *Data Analysis For Social Science.* Princeton University Press. ISBN 9780691199436 (Printed)

   Hands on, includes introduction to R and many examples.

 - Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, M.J.. (2016). *Impact Evaluation in Practice.* 2nd. World Bank Group. ISBN 9781464807794 (Printed)

   Intuitive, practical perspective, many examples from field

   https://openknowledge.worldbank.org/handle/10986/25030

 - Ethan Bueno de Mesquita and Anthony Fowler. (2021). *Thinking Clearly with Data.* Princeton University Press. ISBN 9780691214351 (Printed)

   More conceptual. Covers additional topics.

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.