

Big & Smart Data Analysis STG-MA-B04-23

Administrative Information

Course Title	Big & Smart Data Analysis
Code	STG-MA-B04-23
ECTS	3
Professor/ Instructor	Tommaso Nannicini
E-mail	tommaso.nannicini@eui.eu
Teaching Assistant (TA)	Gonçalo Lima
Email	goncalo.lima@eui.eu
Academic year	2023/2024
Semester	2 nd semester, Spring 2024

Teaching and Learning Information

Rationale	Big data are everywhere, from big firms to big government. Policy leaders must know the use (and misuse) of data, either to regulate them or to use them to design welfare-enhancing policies. Moreover, they must be aware that the enormous predictive power of big data may sometimes mislead us into drawing false conclusion and hamper sound policy making. That's why policy leaders must also know smart data, that is, design-based data analysis meant to distinguish spurious correlations from causal relationships. This seminar will discuss these topics through both theory and case studies, so as to equip students with the essential skills to navigate the data-driven landscape of governance. Nowadays, data analysis cannot be missing in any high-level public policy training, even more so at the transnational level.
Description	In the initial part of this seminar, we will embark on a short discussion of fundamental concepts in data analysis (and data manipulation). We will begin by honing students' ability to discern and critique misleading statistics, ensuring they can identify potential pitfalls and misrepresentations in data-driven arguments. Moving forward, our focus will shift to understanding and

	addressing spurious correlations, delving into the intricacies of endogeneity and omitted variables to uncover the real patterns hidden within data. Finally, we will dive into the intricate task of disclosing causal relationships, emphasizing the importance of exogeneity and desing-based policy evaluation in drawing meaningful conclusions from data. Throughout this course, the discussion will be rigorous but not formal, leveraging intuition more than equations. And we will ground our discussions in practical applications drawn from empirical studies within the social sciences. These case studies will provide real-world context and relevance to the topics covered, including the analysis of politician compensation, the examination of visa status and its correlation with crime, the study of public resource allocation and its impact on corruption, the evaluation of political campaigning strategies and their influence on voting behavior, and so on.		
Objectives	This seminar aims to providing students with a robust (albeit introductory) foundation in data analysis for policy making. By the end of the seminar, students will be able to both spot bad data analysis and understand the pillars of sound data analysis. The aim of the seminar is to enable students to apply data-driven approaches in their future roles within government, research, or policy analysis.		
Learning outcomes (LOs) reflecting the objectives	<p>On successful completion of this seminar, students will:</p> <ul style="list-style-type: none"> • Develop critical thinking skills to discern the reliability of data analyses provided by the media, government agencies, or private entities. • Distinguish between spurious correlations and causal relationships. • Explain the power and potential use of big data. • Analyze and draw evidence-based insights from case studies. 		
Pre-requisite(s)	None.	Co-requisite(s)	None.
Teaching and learning methods	Theoretical principles underpinning data analysis will be presented in the first part of the class, then the second part will delve into a discussion of empirical applications related to these principles. Weekly meetings with the tutor will provide students with additional case studies.		
Status	Compulsory	X	Elective
Assessment	Assessment method(s) and percentage of weighting		
	Students will be asked to write a short essay consisting in a fully designed (although hypothetical) evaluation project. The essay should specify: 1) the causal relationship to be disclosed; 2) the nature of the necessary data; 3) the source of exogenous variation; 4) the statistical methods to be used; 5) the interpretation of (potential) results including (potential) caveats.		
	Specific assessment criteria and grade percentage		
	The above essay will account for 75% of the final grading, class participation and discussion for the remaining 25%.		

	<p>Format requirements and layout of written assignments</p> <p><u>-Paper format:</u> Standard A4 format</p> <p><u>-Margins:</u> Same margins (e.g., 2.5 cm, top/bottom, right/left).</p> <p><u>-Font:</u></p> <ul style="list-style-type: none"> • The text must be in embedded, 12-point character font. • Choose a commonly used font that provides a full character set. • Font requirements apply to all text, including captions, footnotes, citations, etc. • Headings are in bold, and subheadings are in italics. • A smaller font size (10-point character) may be appropriate for footnotes or other material outside of the main text. <p><u>-Line spacing:</u> The document must be 1.5 spaced except for quotations as paragraphs, captions, lists, graphs, charts, footnotes/endnotes, bibliographic entries, items within tables, and lists in appendices. Exceptions may have the following format:</p> <ul style="list-style-type: none"> • Quotations and footnotes may be single-spaced. <p><u>-Pagination:</u> Pages must be numbered.</p>		
<p>Academic Misconduct</p>	<p>During any academic activity, especially but not limited to the completion of assignments and the participation in exams, students are expected to refrain from any form of misconduct as defined by the EUI Code of Ethics in Academic Research. While the EUI Code of Ethics in Academic Research defines “Research Misconduct”, these definitions apply equally to all academic activities within the Master’s programme, including those that are not necessarily research oriented. According to the EUI Code of Ethics, academic misconduct “implies (and is not limited to) fabrication, falsification, plagiarism or deception in proposing, carrying out or reporting results of research and deliberate, dangerous or negligent deviations from accepted practice in carrying out research. [...] Misconduct also includes any plan or attempt to do any of these things.” See also Chapter IV. Misconduct in Academic Research of the Code of Ethics in Academic Research.</p>		
<p>Hours</p>	<p>Total contact hours</p> <p>Lecture: 12 hours Tutor: 6 hours</p>	<p>Total non-contact hours</p> <p>Independent study and essay writing: 57 hours</p>	<p>Total learning hours</p> <p>Total: 75 hours</p>
<p>Attendance</p>	<p>According to the Master’s programme Rules and Regulations, the minimum attendance requirement is 75%.</p>		

Exams information	The final essay must be handed in by March 12, 2024 .
--------------------------	--

Sessions	
Session # 1 Date: 30/01/2024 Time: 14-16 Delivery mode: in person	Topic: How to Lie with Statistics Description: We will discuss how statistical analysis can be used to manipulate the public discourse and advocate flawed policies (for instance, drawing inference from averages in biased samples). We will also introduce basic notions to think about causality and counterfactuals. Useful, but non compulsory, references: <ul style="list-style-type: none"> ✓ Darrell Huff, "How to Lie with Statistics," Penguin, 1991 ✓ Judea Pearl & Dana Mackenzie, "The Book of Why. The New Science of Cause and Effect," Penguin, 2018
Session # 2 Date: 06/02/2024 Time: 14-16 Delivery mode: in person	Topic: Regression Analysis and Spurious Correlations Description: We will discuss the intuition behind basic (bivariate and multivariate) regression analysis and how, again, it can trick us into flawed inference and spurious correlations. Case studies will be discussed. Main reference: <ul style="list-style-type: none"> ✓ Joshua D. Angrist and Jörn-Steffan Pischke, "Mastering 'Metrics,'" Princeton University Press, 2015, Chapters 2 & part of 3
Session # 3 13/02/2024 Time: 14-16 Delivery mode: in person	Topic: Causality, Regressions, and Randomized Trials Description: Building on our previous discussions of regression analysis, this class shifts focus toward scenarios where empirical findings can be interpreted causally. This shift is often facilitated by specific experimental designs or natural experiments, such as instrumental variables. Case studies will be discussed. Main reference: <ul style="list-style-type: none"> ✓ Joshua D. Angrist and Jörn-Steffan Pischke, "Mastering 'Metrics,'" Princeton University Press, 2015, Chapters 1 & part of 3
Session # 4 20/02/2024 Time: 14-16	Topic: Disclosing Causality with Experimental Methods Description: We will discuss how to disclose causal relationships using experimental methods, such as Randomized Controlled Trials (RCTs). As empirical applications we will discuss the two papers in the references.

<p>Delivery mode: in person</p>	<p>References (non compulsory):</p> <ul style="list-style-type: none"> ✓ S. Heller et al., “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago,” Quarterly Journal of Economics, 2017 ✓ V. Galasso et al., “Positive Spillovers from Negative Campaigning,” American Journal of Political Science, 2023
<p>Session # 5</p> <p>27/02/2024</p> <p>Time: 14-16</p> <p>Delivery mode: in person</p>	<p>Topic: Disclosing Causality with Quasi-Experimental Methods</p> <p>Description: We will discuss how to disclose causal relationships using quasi-experimental methods, such as event studies and Regression Discontinuity Designs (RDDs). As empirical applications we will discuss the two papers in the references.</p> <p>References (non compulsory):</p> <ul style="list-style-type: none"> ✓ F. Brollo et al., “The Political Resource Curse,” American Economic Review, 2013 ✓ P. Pinotti, “Clicking on Heaven’s Door: The Effect of Immigrant Legalization on Crime,” American Economic Review, 2017
<p>Session # 6</p> <p>05/03/2024</p> <p>Time: 14-16</p> <p>Delivery mode: in person</p>	<p>Topic: Evidence-Based Policy Making</p> <p>Description: We will discuss how design-based empirical evaluation methods should guide policy making also in the policy design phase. As a paramount case study, we will discuss the PROGRESA program in Mexico.</p> <p>Reference (non compulsory):</p> <ul style="list-style-type: none"> ✓ E. Skoufias, “PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico,” IFPRI Research Paper n. 139, 2005