Lahore University of Management Sciences

SS388: Rethinking Data: Cause and Effect in the Humanities

Summer 2024

Instructor	Saad Imtiaz
Room No.	TBD
Office Hours	MTWR 11:00AM –12:50 PM
Email	saad.imtiaz@lums.edu.pk
Telephone	
Secretary/TA	
TA Office Hours	
Course URL (if any)	

Course Teaching Methodology
• Teaching Methodology: synchronous or asynchronous or a blend of both • Lecture details: Percentage of recorded and live interaction lectures TBD

Course Basics				
Credit Hours	4			
Lecture(s)	Nbr of Lec(s) Per Week	2	Duration	110 minutes
Recitation/Lab (per week)	Nbr of Lec(s) Per Week		Duration	
Tutorial (per week)	Nbr of Lec(s) Per Week		Duration	

Course Distribution	
Core	
Elective	Elective
Open for Student Category	HSS sophomores, juniors and seniors
Close for Student Category	N/A

COURSE DESCRIPTION

Do Harvard graduates earn more because of their degree or because of their family backgrounds? Cause-effect relationships are crucial in many areas of social science. Unfortunately, these relationships are not always directly implied in the data we encounter. This course helps humanities students understand how social scientists uncover these relationships in data using clever techniques. We do this by applying statistical principles in numerous real-world examples from diverse social science fields such as political science, history, psychology, and economics. Using these examples, we will show how researchers address important questions in policy and research situations.

The course begins by reviewing basic statistical techniques, like regression analysis. Then, we'll delve into the fundamentals of experimental design, covering topics like randomized controlled trials (RCTs), natural experiments, and quasi experiments. In the last part of the course, we'll explore the challenges of assessing how precise our estimates are and how that impacts our overall conclusions.

The ability to connect cause and effect is highly valued in applied research, public policy, data science, and the technology industry. As a humanities major, this course will also equip you with the tools to understand and engage with data analyses found in research papers and policy reports. Many of these materials directly apply the techniques covered in the course.

MGSHSS, LUMS and particularly this class, is a harassment free zone. There is absolutely zero tolerance for any behaviour that is intended or has the expected result of making anyone uncomfortable and negatively impacts the class environment, or any individual's ability to work to the best of their potential.

If you think that you may be a victim of harassment, or if you have observed any harassment occurring in the purview of this class, please reach out and speak to me. If you are a victim, I strongly encourage you to reach out to the Office of Accessibility and Inclusion at oai@lums.edu.pk or the sexual harassment inquiry committee at harassment@lums.edu.pk for any queries, clarifications, or advice. You may choose to file an informal or a formal complaint to put an end of offending behavior.

Lahore University of Management Sciences

COURSE PREREQUISITE(S)	
•	Statistics and Data Analysis, or Applies Statistics for Humanities (SS187), or Quantitative Research Methods
•	ANTI-REQUISITE: Impact Evaluation

COURSE OBJECTIVES	
	By the end of the course, student should be able to:
•	1. Acknowledge why causality is important in the humanities
•	2. Understand why looking at secondary data will not give us a causal inference
•	3. Appreciate the importance of experimental/random variation for causality
•	4. Understand how one can do laboratory style experiments in the social sciences
•	5. Recognize when experiments cannot be run, and what to do when that happens
•	6. Be able to distinguish across and implement quasi experimental techniques`

Grading break up: Component Details and weightages
Attendance and CP (10%) Assignments (20%) Quizzes (10%) Mid-term (25%) Final (25%) Presentation (10%)* *Students will be presenting an assigned paper from each module, as a part of a group of 4

Examination Detail	
Midterm Exam	Yes/No: Yes Combine Separate: Duration: 2 hours Exam Specifications: MCQ's, followed by written section
Final Exam	Yes/No: Yes Combine Separate: Duration: 2 hours Exam Specifications: MCQ's, followed by written section

Ethics
The strength of the university depends on academic and personal integrity. Students are expected to abide by the rules of academic and personal honesty. Serious ethical violations include cheating, plagiarism, reuse of essays, improper use of the internet and electronic services, unauthorized collaboration, alteration of graded essays, forgery, lying, and unfair competition. For more information on ethics, please refer to the student handbook and the plagiarism document distributed by the Department of Humanities and Social Sciences.

Textbook(s)/Supplementary Readings
Mostly Harmless Econometrics by Joshua Angrist and John Stephen Pishke Statistics Essentials for dummies by Deborah Rumsey (Chapters 6-9 for inferential statistics) Supplementary reading: Mastering Metrics by Joshua Angrist and John Stephen Pishke Causal Inference – The Mixtape by Scott Cunningham

Lahore University of Management Sciences

COURSE OVERVIEW			
Lecture(s)	**Topics**	**Assigned Readings / Notes**	
1 to 2	A review of inferential statistics	Statistics Essentials Chapter 6-9 Appreciate how a select few can reasonably speak for many using a sample. We will see how a small sample can say a great many things about a population. A revision of random sampling and the central limit theorem. Application of the CLT, including confidence intervals and hypothesis tests. Class example: Is the rate of child marriage higher in Pakistan than in India? How can we be sure the differences are not a product of random chance?	
3 to 4	A review of regression	[Mostly Harmless Econometrics chapter 3.1 and 3.2] We will review how to fit a simple best fit (regression) line and the issues that come up right after. We will also see how this is one of the best ways to get close to causal inference, under the right conditions. We will also see how it more often fails, including cases of omitted variables, reverse causality and selection. Class example: What is the gender wage gap? Can we quantify this in observed data? What about gender wage gap, conditional on demographic attributes? [Assignment 1 on the basics of data analysis and regression on STATA (optionally on R)]	
5	Potential Outcomes	[Mostly Harmless Econometrics chapter 2.1 and 2.2] A brief look at the Rubin Causal Model which is the basis for modern causal inference. We will look at why observed data suffers from selection bias, and how we can correct for it using the random assignment of treatment. Class example: Does attending an elite school like Aitchison College or LUMS result in better outcomes or is it the case that the students with better *potential outcomes* select themselves into elite schools?	
6 to 7	Randomized Experiments (RCTs)	[Mostly Harmless Econometrics chapter 2.3] A look at the design and implementation of randomized experiments (or randomized clinical trials), which are accepted as the gold standard for causal inference (also known as A/B testing in the tech sector). Journal articles covered in class: "Corruption in driving license process in India" By Marianne Bertrand, Simeon Djankov, Rema Hanna, Sendhil Mullainathan (2008) Journal article covered as part of the assignment: "The Demand for, and Impact of, Learning HIV Status." Thorntorn (2008) [Assignment 2: replicating the main results from a published paper that uses randomized experiments]	

7 to 8	Group presentations on RCTs	Each group will be assigned one of the following papers to present on: “Experimental Evidence on Female Voting Behavior in Pakistan” By G. Mansuri and X. Gine (2018) “Are Emily and Greg more employable than Lakisha and Jamal?” by Bertrand and Mullinathan (2004) "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." By Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer (August 2009) Generating skilled self employment in developing countries: experimental evidence from Uganda by Christopher Blattman, Nathan Fiala, & Sebastian Martinez, 2007 “Head or Tails: The impact of a coin toss on major life decisions” by Stephen. Levitt (2016) “Teacher Performance Pay: Experimental Evidence from India” by Muralidharan and Sundaraman, Journal of Political Economy (2014) “The Red Sneakers Effect: Inferring Status and Competence from Signals of Nonconformity” by BELLEZZA et al. (2013)	
9 to 10	Instrumental variables (IV)	[Mostly Harmless Econometrics chapter 4] Introduction to quasi experiments, starting with instrumental variables (IV). Instrumental variables are great at extracting ‘experimental variation’ in observed data when running experiments is not feasible. Journal articles covered in class: “The colonial origins of development” by Acemoglu, Johnson, Robinson, (2001) “The Effects of Rural Electrification on Employment: New Evidence from South Africa” by Taryn Dinkelaman (2010) [Assignment 3: comparing the results from simulated IV data and OLS]	
11 to 12	Group presentations on instrumental variable design	Each group will be assigned one of the following papers to present on: “On the origins of gender roles: women and the plough” by Alberto Alesina, Paola Giuliano, Nathan Nunn, 2013 “Colonialism and modern income- islands as a natural experiment.” by Feyrer and Sacerdote (2009) “Direct versus Indirect Colonial Rule in India- Long-term Consequences.” By Lakshmi Iyer (2008) “The Legacy of Colonial Land Tenure Systems in India.” by Banerjee & Iyer (2005) “The Slave Trade and the Origins of Mistrust in Africa” by Nunn and Wantchekon (2011) “Does Compulsory School Attendance Affect Schooling and Earnings?” Angrist and Krueger (1991)	
13	Midterm review session	A review of concepts covered in the first half of the course	
14	Midterm exam		

15 to 17	Fixed Effects + The difference-in-differences approach (DID)	[Mostly Harmless Econometrics chapter 5.1 and 5.2] A powerful tool to evaluate natural experiments using the simple 'before-after' idea across a treatment and control. DID is the most popular quasi experimental technique for its simplicity, as long as one can keep track key assumptions. Journal Article: "The power of tv: cable television and women's status in India" by R. Jensen & E. Oster (2009) "Minimum wage and employment" by Card and Kreuger (1994) [Assignment 4: Replicate the main results from Card and Kreuger]	
18 to 19	Group presentations on DID	Each group will be assigned one of the following papers to present on: "Transparency and Performance" by Honig, Lall, and Parks (2022) "Soap Operas and Fertility- Evidence from Brazil" by Ferrara et al. (2012) "Twin studies, return to education. Ashenfelter and Alan Krueger (1994) "Wine and phylloxera in 19th century France" by Banerjee et al. (2007) "Curriculum and Ideology" by Cantoni et al. (2017) "Entitled to Work- Urban Property Rights and Labor Supply in Peru" by Erica Field (2007) "The Impact of the Mariel Boatlift on the Miami Labor Market" by David Card (1990)	
20 to 21	Regression Discontinuity design (RDD)	[Mostly Harmless Econometrics chapter 5.2.1 and 6] RDD is the third quasi experimental technique we will cover in this course. Real life situations with arbitrary rules often given rise to 'discontinuities' in treatment assignment. We can exploit these to look at situations which naturally look like an experiment. Class example(s): Does giving free laptops to Pakistani youth improve their education outcomes? What if we compare someone who barely qualified for the laptop with someone who just missed the cut? Does the dean's honors list actually improve job market outcomes? What is we compare a student with 5.99 GPA to one with 5.61? Journal articles covered in class: "The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age" by Christopher Carpenter and Carlos Dobkin (2009) "Randomized experiments from non-random selection in U.S. House elections" by D. Lee (2008) "The effect of aid on growth: evidence from a Quasi-experiment" by Galliiani et al. (2018) [Assignment 5: Replicate the main results from Carpenter and Dobkin]	
22 to 23	Group presentations on RDD	Each group will be assigned one of the following papers to present on:	

		"Effects of School Quality on Student Achievement" by Lucas and Mbiti (2014) "Going to a Better School" by Pop-Eleches and Urquiol (2013) "Home Computer Use and the Development of Human Capital" by Malamud and Pop-Eleches (2011) "Government Transfers and Political Support" by Marco Manacorda (2011)" "Rural Infrastructure Development and Economic Activity" by Chaurey and Le (2019) "Politics and Local Economic Growth in India" by Asher and Novosad (2017)	
24	Issues with measuring precision	[Mostly Harmless Econometrics chapter 8] We have looked extensively at bias in observed data and how to correct for it. Now we turn towards precisions around point estimates, which is hotly debated in modern data science. We will look at how to to correct standard errors: accounting for heteroskedasticity, serial correlation and clustered treatment.	
25 to 26	Modern quasi experimental techniques	We will review the latest updates in the fields of modern quasi experiments including spatial first differences and synthetic controls. Journal article: 'Water Constraints to Agricultural Productivity in Bhutan' Felipe Dizon, Saad Imtiaz, and Jisang Yu (2022) "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitude" by Albarah et al. (2021)	
27	Final review session	A review of concepts covered in the second half of the course	
28	Final exam		

*List of all readings in the course. Papers in **bold** are covered in class. Others are covered as part of the group presentations.*

<u>Randomized experiments:</u>
"The Demand for, and Impact of, Learning HIV Status." Thorntorn (2008)

"Corruption in driving license process in India" By Marianne Bertrand, Simeon Djankov, Rema Hanna, Sendhil Mullainathan (2008)

"Experimental Evidence on Female Voting Behavior in Pakistan" By G. Mansuri and X. Gine (2018)

"Are Emily and Greg more employable than Lakisha and Jamal?"

"Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." By Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer (August 2009)

Generating skilled self employment in developing countries: experimental evidence from Uganda by Christopher Blattman, Nathan Fiala, & Sebastian Martinez, 2007

"Head or Tails: The impact of a coin toss on major life decisions" by Stephen. Levitt (2016)

Lahore University of Management Sciences

"Teacher Performance Pay: Experimental Evidence from India Teacher Performance Pay: Experimental Evidence from India" by Muralidaran and Sundaraman, Journal of Political Economy (2014)

"The Red Sneakers Effect: Inferring Status and Competence from Signals of Nonconformity" by BELLEZZA et al. (2013)

Instrumental Variables:

"The colonial origins of development" by Acemoglu, Johnson, Robinson, (2001)

"The Effects of Rural Electrification on Employment: New Evidence from South Africa" by Taryn Dinkelaman (2010)

"On the origins of gender roles: women and the plough" by Alberto Alesina, Paola Giuliano, Nathan Nunn, 2013

"Colonialism and modern income- islands as a natural experiment." by Feyrer and Sacerdote (2009)

"Direct versus Indirect Colonial Rule in India- Long-term Consequences." By Lakshmi Iyer (2008)

"The Legacy of Colonial Land Tenure Systems in India." by Banerjee & Iyer (2005)

"The Slave Trade and the Origins of Mistrust in Africa" by Nunn and Wantchekon (2011)

"Does Compulsory School Attendance Affect Schooling and Earnings?" Angrist and Krueger (1991)

Difference in Differences

"The power of tv: cable television and women's status in India" by R. Jensen & E. Oster (2009)

"Minimum wage and employment" by Card and Kreuger (1994)

"Transparency and Performance" by Honig, Lall, and Parks (2022)

"Soap Operas and Fertility- Evidence from Brazil" by Ferrara et al. (2012)

"Twin studies, return to education" by Ashenfelter and Alan Krueger (1994)

"Wine and phylloxera in 19th century France" by Banerjee et al. (2007)

"Curriculum and Ideology" by Cantoni et al. (2017)

"Entitled to Work- Urban Property Rights and Labor Supply in Peru" by Erica Field (2007)

"The Impact of the Mariel Boatlift on the Miami Labor Market" by David Card (1990)

Regression Discontinuity Design

"Randomized experiments from non-random selection in U.S. House elections" by D. Lee (2008)

"The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age" by Christopher Carpenter and Carlos Dobkin (2009)

"The effect of aid on growth: evidence from a Quasi-experiment" by Galliiani et al. (2018)

"Effects of School Quality on Student Achievement" by Lucas and Mbiti (2014)

"Going to a Better School" by Pop-Eleches and Urquiol (2013)

"Home Computer Use and the Development of Human Capital" by Malamud and Pop-Eleches (2011)

"Government Transfers and Political Support" by Marco Manacorda (2011)"

Lahore University of Management Sciences

"Rural Infrastructure Development and Economic Activity" by Chaurey and Le (2019)

"Politics and Local Economic Growth in India" by Asher and Novosad (2017)

Spatial First Differences

"Water Constraints to Agricultural Productivity in Bhutan" by Felipe Dizon, Saad Imtiaz, and Jisang Yu (2022)

Synthetic Controls

"Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitude" by Albarah et al. (2021)