

# GOV1368 Section 1

Mohit Karnani

Harvard

Fall 2024

# Agenda

Introduction

Descriptive Statistics

Application

AI

# Contents

Introduction

Descriptive Statistics

Application

AI

# Introduction

Hi! :)

My name is Mohit Karnani, I received my PhD in Economics and Statistics from MIT, and I am a Postdoctoral Researcher at HKS.

I'm originally from Chile, with Indian roots, and I moved to the US in 2018.

I do research on a mix of topics spanning Education, Matching Systems, Economic Development and Causal Inference. —>

How about you?

## Essays on Empirical Matching Systems

by

Mohit Karnani

Submitted to the Department of Economics and the Statistics and Data Science Center  
on December 15, 2023 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

### ABSTRACT

This dissertation is a collection of three papers on empirical methods in matching systems. In Chapter 1, I study the estimation of treatment effects in the context of randomized controlled trials conducted with participants in matching systems. I show how conventional methods fail to account for the interference across outcomes induced by matching systems, and therefore yield invalid estimates of causal parameters. I propose a method that solves the interference problem and apply it in two empirical settings. Chapter 2 studies the relevance of the configuration of on- and off-platform options when centralized matching systems operate alongside a decentralized matching process. In these situations, the existence of off-platform options in a decentralized system can affect the outcomes of participants in the centralized system who seek to be matched to on-platform options. We show this by developing and estimating a structural model that considers the interplay between on- and off-platform options in a matching system. Chapter 3 studies the causal effects of different screening and recruiting policies affecting applicants in the Chilean centralized college match. We show how machine learning methods can enhance these screening and recruiting policies.

## These Sections

- ▶ Sections will focus on statistical methods.
- ▶ Install Stata and bring your computers to sections.
- ▶ Sections will be *very applied*, you will not be tested on hard math, but rather on understanding the methods and how to use them.
- ▶ It will also be very useful to understand these methods for your final project (and probably for your career in general)!
- ▶ We will devote some time to identification (i.e. how these methods can be used to estimate causal effects with large amounts of data), but not much to inference (i.e. how can we quantify the uncertainty of these estimators using finite data).
- ▶ Note: even though this is not a math/stats/methods/metrics class, you really need to understand how these statistical methods work, and how they are used to arrive at the conclusions in the papers we will cover this semester. Just understanding the results/conclusions is not enough.

# Contents

Introduction

Descriptive Statistics

Application

AI

# Statistics Review

- ▶ Throughout the course, we will study *random variables*, both *discrete* (e.g. race indicators) and *continuous* (e.g. test scores).

**notation** {We will usually denote random variables with a capital letter (e.g.  $S$  or  $T$ ), or just the name of the random variable (e.g. *Score* or *Treatment*).}

- ▶ These random variables follow a *distribution* in the *population*, which we can't observe.

**notation** {If a random variable  $X$  follows a distribution  $F$ , we write this as  $X \sim F$ .}

- ▶ What we do observe is *data* from a *sample*.

**notation** {If we have a sample of  $n$  students, each denoted by a subindex  $i \in \{1, \dots, n\}$ , then each test score  $S_i$  is a random variable and we will observe  $n$  realizations in our data.}

- ▶ Statistical inference is about using the data to learn characteristics of the population. For example, we might use a sample covariance between  $\{S_i\}_{i=1}^n$  and  $\{T_i\}_{i=1}^n$  to learn something about a population covariance  $\text{cov}(S, T)$ .
- ▶ *Causal inference* goes a step deeper and focuses on the structure of the population distribution  $F$ , the *why*, not only the *what*; inferring that  $\text{cov}(S, T) > 0$  is not enough, we want to know if  $T$  *causes*  $S$  or vice versa, by how much, why (mechanisms), etc.

# Descriptive Statistics

- ▶ We will start with a very simple, yet powerful tool: descriptive statistics.
- ▶ Descriptive statistics are functions of the data; they use a sample and compute a number that describes a characteristic of the sample.
- ▶ In principle, they can also be used to infer characteristics of the population, but only under certain conditions (e.g. iid sampling).
- ▶ Measures of central tendency: Mean, Median, Mode
- ▶ Measures of dispersion: Variance, Standard Deviation, Range, Interquartile Range
- ▶ Order statistics: Quartiles, Quintiles, Deciles, Percentiles



# Contents

Introduction

Descriptive Statistics

Application

AI

# Application: Attendance in MA

Stata

# Contents

Introduction

Descriptive Statistics

Application

AI

# A quick note on AI

From the syllabus:

*The use of AI services must be acknowledged, and material extracted from this source must be in quotation marks or otherwise precisely identified.*

## A quick note on AI

From the syllabus:

*The use of AI services must be acknowledged, and material extracted from this source must be in quotation marks or otherwise precisely identified.*

From ChatGPT:

*ChatGPT can make mistakes. Check important info.*

## A quick note on AI

From the syllabus:

*The use of AI services must be acknowledged, and material extracted from this source must be in quotation marks or otherwise precisely identified.*

From ChatGPT:

*ChatGPT can make mistakes. Check important info.*

Recommendations:

- ▶ Avoid using AI for submissions (definitely do not copy-paste).
- ▶ Use it as a study companion if you want, but remember AI can make mistakes. For example, if you want to use AI to summarize a paper, definitely read the paper, and then maybe compare notes. If you don't read the paper and just rely on the summary, there is a high chance you will learn from "hallucinations".
- ▶ It's fair game to use AI to debug your code, e.g. for your final project. This is not a CS class and you won't be evaluated on these skills. However, you should still be in command, and AI should just be your copilot.

ChatGPT: <https://www.aeaweb.org/articles?id=10.1257/aer.20131528>

write an implementation of the algorithm used in the Chilean college admission system



The Chilean college admission system uses a variant of the Deferred Acceptance algorithm called the "Rural Deferred Acceptance with Ties" (RDAT) algorithm. This algorithm is used to assign students to universities based on their academic performance and other factors, such as geographic location, socioeconomic status, and whether they come from a rural or urban area.

ChatGPT: <https://www.aeaweb.org/articles?id=10.1257/aer.20131528>

write an implementation of the algorithm used in the Chilean college admission system



The Chilean college admission system uses a variant of the Deferred Acceptance algorithm called the "Rural Deferred Acceptance with Ties" (RDAT) algorithm. This algorithm is used to assign students to universities based on their academic performance and other factors, such as geographic location, socioeconomic status, and whether they come from a rural or urban area.

Can you give me your source for this claim?: "The Chilean college admission system uses a variant of the Deferred Acceptance algorithm called the "Rural Deferred Acceptance with Ties" (RDAT) algorithm."



Sure, my source for this claim is a research article titled "The Chilean College Admissions System: Matching, Incentives, and Constraints" by Alvin E. Roth, Axel Ockenfels, and Max H. Bazerman, which was published in the American Economic Review in 2015. In the paper, the authors describe the Chilean college admissions system and its use of the Rural Deferred Acceptance with Ties (RDAT) algorithm. They also provide a detailed explanation of the algorithm and its properties. The paper can be accessed here: <https://www.aeaweb.org/articles?id=10.1257/aer.20131528>.