

Problem Set 3: Causal Inference

Due: 5pm on Friday, February 14.

Note

Want to see how you'll be evaluated? Check out the [rubric](#)

Student identifier: [type your anonymous identifier here]

- Use this [.qmd template](#) to complete the problem set
- In Canvas, you will upload the PDF produced by your .qmd file
- Put your identifier above, not your name! We want anonymous grading to be possible

This problem set is based on:

Bertrand, M & Mullainathan, S. 2004. “[Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.](#)” American Economic Review 94(4):991–1013.

Read the first 10 pages of the paper (through the end of section 2). In this paper,

- the unit of analysis is a resume submitted to a job opening
- the treatment is the name at the top of the resume
- the outcome is whether the employer called or emailed back for an interview

Analyzing the experimental data (25 points)

Load packages that our code will use.

```
library(tidyverse)
library(haven)
```

Download the study's data from OpenICPSR: <https://www.openicpsr.org/openicpsr/project/116023/version/V1/view>. This will require creating an account and agreeing to terms for using the data ethically. Put the data in the folder on your computer where this .Rmd is located. Read the data into R using `read_dta`.

```
#d <- read_dta("lakisha_aer.dta")
```

If you have an error, you might need to set your working directory first. This tells R where to look for data files. At the top of RStudio, click Session -> Set Working Directory -> To Source File Location.

You will now see `d` in your Global Environment at the top right of RStudio.

We will use two variables:

Name	Role	Values
<code>call</code>	outcome	1 if resume submission yielded a callback 0 if not
<code>race</code>	category of treatments	<code>b</code> if first name signals Black <code>w</code> if first name signals white

The top of Table 1 reports callback rates: 9.65% for white names and 6.45% for Black names. Reproduce those numbers. Write code that reproduces these numbers.

Causal inference concepts

2.1. (5 points) Fundamental problem

One submitted resume had the name “Emily Baker.” It yielded a callback. The same resume could have had the name “Lakisha Washington.” Explain how the Fundamental Problem of Causal Inference applies to this case (1–2 sentences).

2.2. (5 points) Potential outcomes. Using math, write the following in potential outcomes notation: resume submission 5 would receive a callback if it signaled the name “Emily Baker”. (There are several ways to write a correct answer.) You can either type math or include a picture of handwritten math. See the bottom of this page for help.

2.3. (5 points) Exchangeability

In a sentence, what is the exchangeability assumption in this study? For concreteness, for this question you may suppose that the only names in the study were “Emily Baker” and “Lakisha Washington.” Be sure to explicitly state the treatment and the potential outcomes.

2.4. (5 points) Observational study

Suppose that instead of randomly assigning names to fictitious resumes, the authors had instead analyzed real job applications by people named “Emily Baker” and “Lakisha Washington.” Use mathematical notation with a conditional probability $P(A \mid B)$ to state the following: the probability of being called back was higher among resumes from Emily Baker than among resumes from Lakisha Washington.

2.5 (5 points) Exchangeability violated

The descriptive estimand in the observational study (2.4) is different than the causal estimand in the experimental study (2.2). Suppose the researchers wanted to use the observational study to learn about the quantity in (2.2). Explain one way the exchangeability assumption would be violated.

How to type math

There are two ways to type math within your .qmd file, outside of a code chunk.

- To type in-line math surround your math in a single $$ at each side. Typing $X = 1$ will produce $X = 1$.
- To type math that goes on its own equation line, use $$ before and after the math. Typing $X = 1$ will produce

$$X = 1$$

When typing math, there are a few things you will want to know.

- $$ indicates the start of a subscript: Y_i becomes Y_i
- $$ indicates the start of a superscript: Y^a becomes Y^a
- $$ will let you use several characters in a subscript or superscript: $Y_{unit}^{treatment}$ becomes $Y_{unit}^{treatment}$
- $$ is the vertical conditioning bar: $E(Y \mid X)$ becomes $E(Y \mid X)$.

How to include hand-written math

You are also welcome to handwrite math and take a picture of it. Put your picture in the folder where your .qmd document is located. To include the picture in your writeup, type

`![] (NameOfYourPictureFile.png)`