

## Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview?

### Part V: Focus on Black Applicants

Let's continue working with the data from the experiment in Milwaukee where researchers randomly assigned whether the job applicant had a criminal record. As a reminder, Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is individual job applications.

variable	description
<i>job_id</i>	identifying number of job opening
<i>criminal</i>	whether the job applicant presented himself as having a criminal record (1=yes, 0=no)
<i>race</i>	race of applicant (black or white)
<i>call</i>	whether job application received a call back for a job interview (1=yes, 0=no)

Table 1: Variables in "applications.csv"

In this problem set, we practice answering the following four questions related to causal studies: (1) What is the estimated average treatment effect? (2) Is the effect statistically significant at the 5% level? (3) Can we interpret the effect as causal? And (4) Can we generalize the results?

As always, we start by loading and looking at the data:

```
## load and look at the data
applications <- read.csv("applications.csv") # reads and stores data
head(applications) # shows first observations
##   job_id criminal  race call
## 1      1         0 white    1
## 2      1         1 white    1
## 3      2         1 white    0
## 4      2         0 white    0
## 5      3         1 white    0
## 6      3         0 white    0
```

Let's now focus on the pair of black applicants. To do so, we can run the piece of code below, which creates a new dataframe containing only the job applications that correspond to the black applicants. (It uses the `[]` operator to extract a selection of observations from a dataframe, as explained on page 208 of DSS.)

```
## create new dataframe containing only the job applications for black applicants
applications_black <- applications[applications$race=="black", ]
```

Now, we are ready to start our analysis:

1. In the dataset about the black applicants, what is the estimated average causal effect of having a criminal record on the probability of receiving a call back for a job interview?

- a. Fit a linear model to the data in such a way that the estimated slope coefficient is equivalent to the difference-in-means estimator you are interested in and store the fitted model in an object called *fit* (R code only). (5 points)
  - b. What is the estimated slope coefficient,  $\hat{\beta}$ ? (5 points)
  - c. Now, let's answer the question: What is the estimated average treatment effect? Provide a full substantive answer (make sure to include the assumption, why the assumption is reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (10 points)
2. Is the effect statistically significant at the 5% level?
- a. Let's start by specifying the null and alternative hypotheses. Please provide both the mathematical notations and their meaning. (2.5 points)
  - b. What is the value of the observed test statistic,  $z^{obs}$ ? (Hint: the code `summary()$coeff` might be helpful here.) (2.5 points)
  - c. What is the associated p-value? (2.5 points)
  - d. Now, let's answer the question: Is the effect statistically significant at the 5% level? Please provide your reasoning. (2.5 points)
3. Can we interpret the estimated effect as causal? In other words, how strong is the internal validity of this study? Have the researchers accurately measured the average causal effect on the sample of candidates who were part of the study? Please explain your reasoning. (10 points)
4. Can we generalize the results? In other words, how strong is the external validity of this study? Please explain your reasoning and be specific about what population you think the findings can or cannot be generalized to. (10 points)