

Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? Part III: Estimating an Average Causal Effect (with Solutions)

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 (1) how to estimate an average treatment effect using data from a randomized experiment and (2) how to write a conclusion statement.

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
```

To simplify our analysis, let's focus on one of the two pairs: the pair of white 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 white 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 white applicants
applications_white <- applications[applications$race=="white", ]
```

Now, we are ready to start our analysis:

1. Considering that the dataset we are analyzing comes from a randomized experiment, what can we compute to estimate the average causal effect of having a criminal record on the number of new (or repaired) drinking call facilities? Please provide the name of the estimator. (5 points)

Answer: The difference-in-means estimator.

2. In the dataset about the white applicants, what is the probability of receiving a call back for a job interview among the job applications with a criminal record? Please answer with a full sentence. (10 points)

R code:

```
## compute mean of call among job applications with criminal records  
mean(applications_white$call [ applications_white$criminal==1])  
## [1] 0.1666667
```

(Recall: we use `[]` to subset a variable; inside the square brackets, we specify the criterion of selection. For example, we can use the relational operator `==` to set a logical test; only the observations for which the logical test are true will be extracted. In the case above, we extract from *call* the observations that belong to the job applications that were assigned to have a criminal record (for which *criminal* equals 1) and then calculate the mean of those observations.)

Answer: In the dataset about the white applicants, the probability of receiving a call back for a job interview among the job applications with a criminal record is of 17%. (Note: The mean of a binary variable should be interpreted as the proportion (or probability) of the observations that have the characteristic identified by the variable. Here, *call* is binary and identifies the job applications that received a call back for a job interview. The unit of measurement is %, after multiplying the output by 100. Here: $0.17 \times 100 = 17\%$.)

3. In the dataset about the white applicants, what is the probability of receiving a call back for a job interview among the job applications without a criminal record? Please answer with a full sentence. (10 points)

R code:

```
## compute mean of call among job applications without criminal records  
mean(applications_white$call [ applications_white$criminal==0])  
## [1] 0.34
```

Answer: In the dataset about the white applicants, the probability of receiving a call back for a job interview among the job applications without a criminal record is of 34%.

4. What is the estimated average causal effect of having a criminal record on the probability of receiving a call back for a job interview? Please 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) (25 points)

R code:

```
## compute the difference in means estimator  
mean(applications_white$call [ applications_white$criminal==1]) -  
  mean(applications_white$call [ applications_white$criminal==0])  
## [1] -0.1733333
```

Answer: Let's start by figuring out each key element separately.

- *What's the assumption?* We assume that the job applications that were assigned to have a criminal record (the treatment group) are comparable to the job applications that were assigned to NOT have a criminal record (the control group). (Note: If this assumption were not true the difference-in-means estimator would NOT produce a valid estimate of the average treatment effect.)
- *Why is the assumption reasonable?* Because the criminal records were assigned at random OR because the data come from a randomized experiment. (Recall: Random treatment assignment makes the treatment and control groups on average identical to each other in all observed and unobserved pre-treatment characteristics.)
- *What's the treatment?* Having a criminal record.
- *What's the outcome?* Probability of receiving a call back for a job interview. (Note: When the outcome variable is binary, we speak of the effect of the treatment on the *probability of the outcome occurring*.)
- *What's the direction, size, and unit of measurement of the average causal effect?* A decrease of 17 percentage points, on average. (Note: It is a decrease because we are measuring change—the change in the outcome variable caused by the treatment—and the difference-in-means estimator is negative. The difference-in-means estimator = proportion of call backs among job applications with a criminal record - proportion of call backs among job applications without a criminal record = 17% - 34% = -17 percentage points. Recall: percentage points is the unit of measurement for the arithmetic difference between two percentages.)

Full answer: Assuming that the job applications that were randomly assigned to have a criminal record were comparable to the job applications that were randomly assigned to NOT have a criminal record (a reasonable assumption since the criminal records were assigned at random), we estimate that having a criminal record decreases the probability of receiving a call back for a job interview by 17 percentage points, on average.

(Note: the following statement is not the correct answer because it doesn't make a causal claim; it only makes an observation: Assuming that the job applications that were randomly assigned to have a criminal record were comparable to the job applications that were randomly assigned to NOT have a criminal record (a reasonable assumption since the criminal records were assigned at random), we estimate that job applications with a criminal record have a probability of receiving a call back for a job interview that it is, on average, 17 percentage points lower than the probability of receiving a call back for a job interview among job applications without a criminal record.)