Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? Part II: Computing and Interpreting Means

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
job_id	identifying number of job opening
criminal	whether the job applicant presented himself as having a criminal record (1=yes, 0=no)
race	race of applicant (black or white)
call	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 how to compute and interpret means, among other things.

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. Use the function dim() to find how many observations are in the original dataframe *applications* and how many observations are in the new dataframe *applications_white*. Provide a full sentence with what you learn from computing these two numbers. (10 points)

2. Use the function mean() to calculate the average of the variable *criminal* in the dataframe *applications_white*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)
3. Use the function mean() to calculate the average of the variable *call* in the dataframe *applications_white*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)
4. If we wanted to estimate the average causal effect of having a criminal record on the probability of getting a call back for a job interview for the white applicants: (10 points)
    a. What would be the treatment variable? Please just provide the name of the variable in the *applications_white* dataframe
    b. What would be the outcome variable? Please just provide the name of the variable in the *applications_white* dataframe
5. In this analysis: (10 points)
    a. What would be the treatment group?
    b. What would be the control group?