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

R code:

```
# compute dimensions of dataframe: rows, columns
dim( applications ) # of original dataframe
## [1] 696   4
dim( applications_white) # of new dataframe
## [1] 300   4
```

Answer: Based on the outputs above, we learn that out of the 696 job applications in the experiment, 300 were submitted by the white applicants.

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)

R code:

```
mean(applications_white$criminal) # calculates the average of criminal
## [1] 0.5
```

(Recall: We use $ to access a variable within a dataframe. To its left, we specify the name of the object where the dataframe is stored, applications_white in this case; to its right, we specify the name of the variable, criminal in this case. Also, we do not use quotes around the names of functions, the names of objects, or the names of elements within an object such as variables.)

Answer: Exactly 50% of the job applications submitted by the white applicants were randomly assigned to have a criminal record, which is consistent with the notion that each tester served in the criminal record condition for an equal number of employment applications. (Note: The word "average" does not appear in the interpretation of this mean because it should be interpreted as a proportion. Recall: The mean of a binary variable should be interpreted as the proportion of the observations that have the characteristic identified by the variable. Here, *criminal* is binary and identifies the job applications that were assigned to have a criminal record. The unit of measurement is %, after multiplying the output by 100. Here: $0.50 \times 100 = 50\%$.)

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)

R code:

```
mean(applications_white$call) # calculates the average of water
## [1] 0.2533333
```

Answer: About 25% of the job applications submitted by the white applicants received a call back for a job interview. (Note: See previous explanation for more details. Here, *call* is also 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: $0.25 \times 100 = 25\%$.)

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

   <u>Answer</u>: (a) *criminal*, (b) *call*.

5. In this analysis: (10 points)

   a. What would be the treatment group?

   b. What would be the control group?

   <u>Answer</u>: (a) the job applications submitted by the white applicants that were randomly assigned to have a criminal record, (b) the job applications submitted by the white applicants that were randomly assigned to NOT have a criminal record. (Note: the unit of observation is job applications so both of these groups are composed of job applications; the treatment group are the job applications that were assigned to receive the treatment—having a criminal record; the control group are the job applications that were assigned to NOT receive the treatment.)