Does Having a Criminal Record Affect the Chances of Receiving A Call Back for a Job Interview? Part IV: Focus on White Applicants (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 (1) how to estimate an average treatment effect using data from a randomized experiment and (2) how to determine whether the estimated average treatment effect is statistically significant at the 5% level.

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. In the dataset about the white 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)

R code:

```
## fit and store linear model
fit <- lm(applications_white$call ~ applications_white$criminal) # or
fit <- lm(call ~ criminal, data = applications_white)
```

(Note: Remember that the lm() function requires an argument of the form Y∼X. Here, *call* is the outcome variable, Y, and *criminal* is the treatment variable, X.)

b. What is the estimated slope coefficient, $\widehat{\beta}$? (5 points)

R code:

```
fit # shows contents of object
##
## Call:
## lm(formula = call ~ criminal, data = applications_white)
##
## Coefficients:
## (Intercept)     criminal
##      0.3400      -0.1733
```

Answer: The estimated slope coefficient, $\widehat{\beta}$, is -0.17.

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)

Answer: 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

among white applicants. (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 is measured in percentage points (after we multiply the output by 100) because the outcome *call* is binary. As a result, both the average outcome for the treatment group and the average outcome for the control group will be measured in percentages, after multiplying the outputs by 100, and 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 among white applicants by 17 percentage points, on average.

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. (10 points)

   Answer: The null and alternative hypotheses are:

   H_0: $\beta=0$ (meaning: having a criminal record has no average causal effect on the probability of receiving a call back for a job interview among white applicants at the population level).

   H_1: $\beta\neq0$ (meaning: having a criminal politician, on average, either increases or decreases the probability of receiving a call back for a job interview among white applicants at the population level)

   (Note that the null and alternative hypotheses refer to β, which is the true average causal effect at the population level, not to $\widehat{\beta}$, which is the estimated average causal effect at the sample level.)

   b. What is the value of the observed test statistic, z^{obs}? (Hint: the code summary()$coeff might be helpful here.) (5 points)

   R code:

```
summary(fit)$coeff
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)  0.3400000 0.03491547  9.737803 1.233829e-e19
## criminal    -0.1733333 0.04937793 -3.510340 5.167157e-e04
```

Answer: The value of the observed test statistic, z^{obs}, is -3.51. (Note: The observed test statistic for regression coefficients equals $\widehat{\beta}$ divided by the standard error of $\widehat{\beta}$. Here, it equals -0.17333/0.04938= -3.51, which is exactly what R provides as the t-value for the coefficient affecting *criminal*, that is, the value in the cell in the second row, third column of the table above.)

c. What is the associated p-value? (5 points)

Answer: The associated p-value is 5.17e-04, which is $5.17 \times 10^{-4} \approx 0.000517$. (Note: We can interpret this as indicating that, if the null hypothesis were true, the probability of observing a test statistic equal to or larger than -3.51 (in absolute value) is about 0.05%. This is a small probability, well below 5%, so we will reject the null hypothesis.)

d. Now, let's answer the question: Is the effect statistically significant at the 5% level? Please provide your reasoning. (10 points)

Answer: Yes, the effect is statistically significant at the 5% level. Because (a) the absolute value of the observed test statistic is greater than 1.96 (|-3.51|>1.96), and/or (b) the p-value is smaller than 0.05 (0.0005<0.05), we reject the null hypothesis and conclude that there is likely to be an average treatment effect different than zero at the population level. In other words, we conclude that having a criminal record is likely to have an average effect different than zero on the probability of receiving a call back for a job interview among white applicants *at the population level*. (Note: You do not need to provide both reasons, (a) and (b). One of them suffices since both procedures should lead to the same conclusion.)