Is There Racial Discrimination in the Labor Market? (with Solutions)

(Based on Marianne Bertrand and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review*, 94 (4): 991–1013.)

Let's answer this question by using data from a randomized experiment where researchers sent out resumes of fictitious candidates to apply for jobs advertised in Boston and Chicago newspapers from mid 2001 to mid 2002. In this experiment, the presumed race of the applicant was randomly assigned. Some of the resumes were given names common among African-Americans (the treatment group) and others were given names common among Whites (the control group).

The dataset is in a file called *resumes.csv*. Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is fictitious job applicants.

variable	description
resume_id	identifying number of each resume
firstname	first name of fictitious job applicant
sex	presumed sex of applicant based on name (female or male)
race	presumed race of applicant based on name (black or white)
call	whether candidate received a call back for a job interview (1=yes, 0=no)

Table 1: Variables in "resumes.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, let's start by loading and looking at the data:

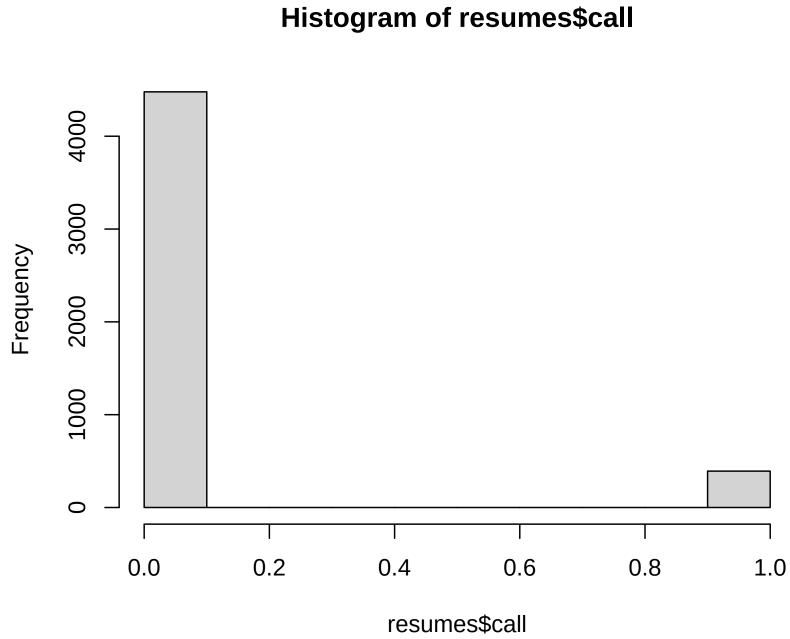
```
## load and look at the data
resumes <- read.csv("resumes.csv") # reads and stores data
head(resumes) # shows first observations
##   resume_id firstname    sex  race call
## 1         1   Allison female white    0
## 2         2   Kristen female white    0
## 3         3   Lakisha female black    0
## 4         4   Latonya female black    0
## 5         5    Carrie female white    0
## 6         6       Jay   male white    0
```

1. What is the estimated average causal effect of having a black sounding name on the probability of being called back for a job interview?

   a. Given our research question, what should be our outcome variable (Y)? Visualize its distribution and comment on the proportion of candidates who were called back for a job interview (i.e., is the proportion large, small, or very small?).

   R code:

```
hist(resumes$call) # creates histogram
```

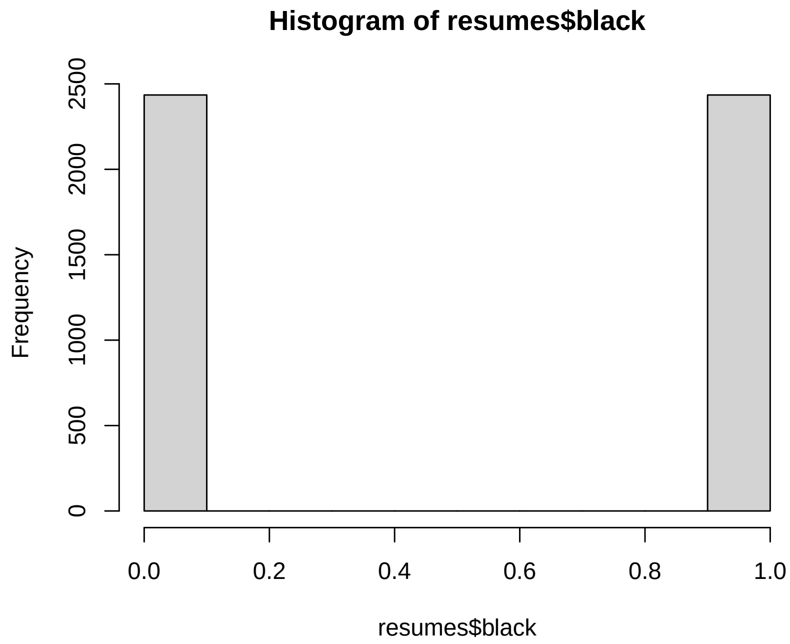


   Answer: The outcome variable should be *call* since that is the variable that identifies the fictitious job candidates who were called back for a job interview. Based on the histogram of *call* above, a very small proportion of candidates were called back for an interview. (Note: To visualize the distribution of one variable, we can create a histogram. The function to create a histogram in R is hist(). The only required argument is the code identifying the variable you want to create the histogram of, which is resumes$call, in this case. In the histogram, we can see that the number of candidates who received a call back (i.e., have a value of 1 in *call*) is below 500, while the number of candidates who did not (i.e., have a value of 0 in *call*) is well above 4,000. Hence, the proportion of candidates who were called back for an interview is very small. In fact, if we were to calculate the proportion directly by running prop.table(table(resumes$call)), we would see that only 8% of candidates were called back. Remember, this is real data! So, when you are applying for jobs, do not be discouraged if you only hear from 2 out of 25 applications you send.)

b. Given our research question, a binary variable identifying the presumed race of the fictitious job candidate should be our treatment variable (X). Since we do not have it in the dataset, we need to create it. Create this variable, name it *black*, and store it inside the dataframe *resumes*. (Hint: the function ifselse() might be a useful here). Visualize its distribution and comment on the number of resumes assigned to treatment and control groups.

R code:

```
resumes$black <- ifelse (resumes$race=="black", 1, 0) # creates new variable
hist (resumes$black) # creates histogram
```



Answer: Based on the histogram of *black* above, more or less the same number of fictitious job candidates were assigned to the treatment group (i.e., given a "black" sounding name) as to the control group (i.e., given a "white" sounding name). (Note: Recall that the function ifelse() requires three arguments and in this order: (1) the logical test (specified using == in this class), (2) the return value if test is false, (3) the return value if test is true. The existing variable we use to create this new binary variable is *race* since that is the variable that identifies treatment. When the value of *race* equals "black", we want the new variable *black* to equal 1; and to equal 0 otherwise. Hence, we can use ifelse(resumes$race=="black", 1, 0) to produce the values of this new variable. Note, however, that since there are only two possible presumed races, we would create the same values if we ran ifelse(resumes$race=="white", 0, 1) instead. Now, to store these values as a variable named *black* inside the dataframe *resumes*, we use the assignment operator <- and specify resumes$black to the left of the assignment operator. Without resumes$ in front of black, you would be creating a variable outside the dataframe, as a new object, which is not what we want.)

c. Now that we have both our Y and X variables, 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).

R code:

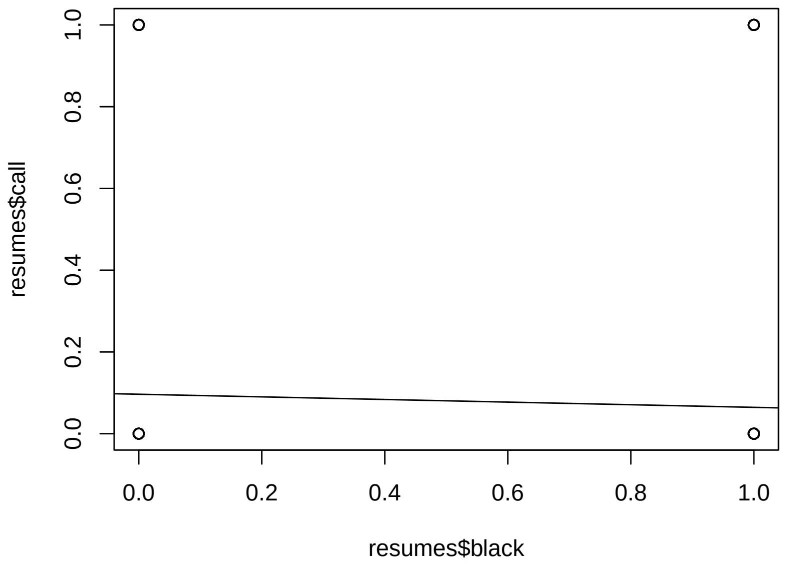
```
fit <- lm(resumes$call ~ resumes$black) # fits and stores linear model
```

(Note: Remember that to fit a linear model in R we use the lm() function. This function requires an argument of the type Y ~ X. Here, *call* is the outcome variable, Y, and *black* is the treatment variable, X. To specify the dataframe where the variables are stored, we can use either the $ operator (as in the code above) or the optional argument data. If we wanted to use the latter, the code to fit the linear model would be lm(call ~ black, data = resumes). Now, to store the fitted linear model as an object named *fit*, we use the assignment operator <- and specify *fit* to its left.)

d. Create a visualization of the relationship between X and Y and add the fitted line. (Hint: The functions plot() and abline() might be helpful here.)

R code:

```
plot(resumes$black, resumes$call) # creates scatter plot
abline(fit) # adds straight line
```



(Note: To visualize the relationship between X and Y, we create a scatter plot with the function plot(). This function requires two arguments and in this order: (1) the X variable, (2) the Y variable. We always plot the X variable along the X axes, and the Y variable along the Y axes. Here, since *black* is the X variable and *call* is the Y variable, the code to create the scatter plot is plot(resumes$black, resumes$call). Alternatively, if we do not want the order of the arguments to matter, we could specify the names of the arguments, x and y, in the code. For example, plot(x=resumes$black, y=resumes$call) and plot(y=resumes$call, x=resumes$black) would produce the same scatter plot as the one above. Now, to add the fitted line to the scatter plot, we use the function abline(). The only required argument is the name of the object where we have stored the fitted line, fit in this case.)

e. What is the estimated slope coefficient, $\widehat{\beta}$?

R code:

```
fit # shows contents of object
##
## Call:
## lm(formula = resumes$call ~ resumes$black)
##
## Coefficients:
##  (Intercept)  resumes$black
##      0.09651       -0.03203
```

Answer: The estimated slope coefficient, $\widehat{\beta}$, is -0.03.

f. 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)

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

- *What's the assumption?* We assume that the fictitious job candidates with black sounding names (the treatment group) are comparable to the fictitious job candidates with white sounding names (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 presumed race of the candidate was 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 black sounding name.

- *What's the outcome?* Probability of being called back for a job interview.

- *What's the direction, size, and unit of measurement of the average causal effect?* A decrease of 3 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 $\widehat{\beta}$, which is equivalent here to the difference-in-means estimator, is negative. Recall that because X is the treatment variable and Y is the outcome variable, $\widehat{\beta}$ is equivalent to the difference-in-means estimator. The size and unit of measurement is 3 percentage points. Recall that when the outcome variable is binary, we should interpret the difference-in-means estimator in percentage points, after multiplying the result by 100. Here, $0.03 \times 100=3$

percentage points. To better understand this, let's look into the details. In this case, the difference-in-means estimator = probability of being called back for job interview for candidates with black sounding names - probability of being called back for job interview for candidates with white sounding names. We can calculate the first term by running `mean(resumes$call[resumes$black==1])` and the second term by running `mean(resumes$call[resumes$black==0])`. (Recall that the mean of a binary variable computes the proportion of 1s, that is, the probability having the characteristic identified by the variable, after multiplying the result by 100.) Looking at the outputs, we find out that the probability of being called back for a job interview is of 6.5% for candidates with black sounding names and of 9.7% for candidates with white sounding names. So, the difference-in-means estimator here is 6.5% - 9.7% = -3.2 percentage points. Recall that percentage point is the unit of measurement of the arithmetic difference between two percentages.)

Full answer: Assuming that the fictitious job candidates with black sounding names were comparable to the fictitious job candidates with white sounding names (a reasonable assumption since the presumed race of the candidate was assigned at random), we estimate that having a black sounding name decreases the probability of being called back for a job interview by 3 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.

      Answer: The null and alternative hypotheses are:

      H_0: β=0 (meaning: having a black sounding name has no average causal effect on the probability of getting a call back at the population level).

      H_1: $\beta \neq 0$ (meaning: having a black sounding name either increases or decreases the probability of getting a call back, on average, 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.)

      R code:

```
summary(fit)$coeff
##                  Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept)    0.09650924 0.005504805 17.531819 8.891552e-67
## resumes$black -0.03203285 0.007784969 -4.114705 3.940803e-05
```

Answer: The value of the observed test statistic, z^{obs}, is -4.11. (Note: The observed test statistic for regression coefficients equals $\widehat{\beta}$ divided by the standard error of $\widehat{\beta}$. Here, it equals -0.03203/0.00778= -4.11, which is exactly what R provides as the t-value for the coefficient affecting *black*, that is, the value in the cell in the second row, third column of the table above.)

c. What is the associated p-value?

Answer: The associated p-value is 3.94e-05, which is $3.94 \times 10^{-5} \approx 0.0000394$. (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 -4.11 (in absolute value) is about 0.004% (0.0000394x $\times$ 100=0.004%). This is a tiny 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.

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 ($|-4.11|>1.96$), and/or (b) the p-value is smaller than 0.05 ($0.00004<0.05$), 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 black sounding name is likely to have an average effect different than zero on the probability of being called back for a job interview *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.)

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.

Answer: Yes, we can interpret the estimated effect as causal. The internal validity of this study is strong because the treatment (having a black sounding name) was assigned at random. Random treatment assignment should have eliminated all confounding variables, making the fictitious job candidates who were given a black sounding name (the treatment group) and the fictitious job candidates who were given a white sounding name (the control group) comparable; that is, they should have the same observed and unobserved characteristics, on average, with the exception of their presumed race. As a result, the difference-in-means estimator should provide a valid estimate of the average causal effect among the candidates in the study. In other words, since the only systematic difference between the treatment and control groups is the presumed race, we can conclude that the observed difference in the outcome (the 3 percentage point decrease in the probability of getting a call back for a job interview) is the direct result of the treatment (having a black sounding name). (Note: To explore whether randomization succeeded in making treatment and control groups comparable, we could explore whether each group has the same proportion of female names. If we run `prop.table(table(resumes$race, resumes$sex), margin=1)`, we learn that indeed the two groups are comparable when it comes to proportion of female names. In both groups, about 76-77% of the names are female.)

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.

   Answer: The answer depends on (i) whether the treatment used in the study is representative of the real-world treatment for which we want to generalize the results, and (ii) whether the sample of observations in the study is representative of the population to which we want to generalize the results. In regards to the first point, the treatment in the study (having a black-sounding name) was observed in the real-world, so it is realistic and comparable to the treatment we want to generalize the result. In regards to the second point, as we learned in the introduction, the experiment was conducted in Boston and Chicago in the early 2000s. Assuming that the sample of job candidates and the sample of job openings involved were representative of these two markets at the time, we could generalize the results to these two cities and conclude that there was racial discrimination in the job market both in Boston and Chicago in the early 2000s. The external validity in this case would be strong. However, if we intended to generalize the results to the whole of the United States, then the sample of job candidates and the sample of job openings involved would likely not be representative and the external validity of this study would be relatively weak. The degree of racial discrimination in the job market in the whole of the U.S. might have been higher or lower than that in Boston and Chicago. To determine it, we would have to conduct another study, where the samples are representative of the target population.