

Resume Experiment Analysis

How much harder is it to get a job in the United States if you are Black than if you are White? Or, expressed differently, what is the *effect* of race on the difficulty of getting a job in the US?

In this exercise, we will be analyzing data from a real world experiment designed to help answer this question. Namely, we will be analyzing data from a randomized experiment in which 4,870 fictitious resumes were sent out to employers in response to job adverts in Boston and Chicago in 2001. The resumes differ in various attributes including the names of the applicants, and different resumes were randomly allocated to job openings.

The “experiment” part of the experiment is that resumes were randomly assigned Black- or White-sounding names, and then watched to see whether employers called the “applicants” with Black-sounding names at the same rate as the applicants with the White-sounding names.

(Which names constituted “Black-sounding names” and “White-sounding names” was determined by analyzing names on Massachusetts birth certificates to determine which names were most associated with Black and White children, and then surveys were used to validate that the names were perceived as being associated with individuals of one racial category or the other).

You can get access to original article here [<https://www.aeaweb.org/articles?id=10.1257/0002828042002561>].

Note to Duke students: if you are on the Duke campus network, you’ll be able to access almost any academic journal articles directly; if you are off campus and want access, you can just go to the Duke Library [<https://library.duke.edu/>] website

and search for the article title. Once you find it, you'll be asked to log in, after which you'll have full access to the article. You will also find this pattern holds true at nearly any major University in the US.

- Download the data set `resume_experiment.dta` from github here [https://github.com/nickeubank/MIDS_Data/tree/master/resume_experiment], or by doing to [www.github.com/nickeubank/MIDS_Data](https://github.com/nickeubank/MIDS_Data) and opening the `resume_experiment` folder.
- For python users, use `read_stata` in pandas to load the data set; For R users, use `read_dta` in haven to load the data set
- `black` is the treatment variable in the data set (whether the resume has a Black-sounding name).
- `call` is the dependent variable of interest (did the employer call the fictitious applicant for an interview)

In addition, the data include a number of variables to describe the other features in each fictitious resume, including applicants education level (`education`), years of experience (`yearsexp`), gender (`female`), computer skills (`computerskills`), and number of previous jobs (`ofjobs`). Each resume has a random selection of these attributes, so on average the Black-named fictitious applicant resumes have the same qualifications as the White-named applicant resumes.

Checking for Balance

The first step in analyzing any experiment is to check whether you have *balance* across your treatment arms—that is to say, do the people who were randomly assigned to the treatment group look like the people who were randomly assigned to the control group. Or in this case, do the resumes that ended up with Black-sounding names look like the resumes with White-sounding names.

Checking for balance is critical for two reasons. First, it's always possible that

random assignment will create profoundly different groups—the *Large of Large Numbers* is only a “law” in the limit. So we want to make sure we have reasonably similar groups from the outset. And second, it’s also always possible that the randomization wasn’t actually implemented correctly—you would be amazed at the number of ways that “random assignment” can go wrong! So if you ever do find you’re getting unbalanced data, you should worry not only about whether the groups have baseline differences, but also whether the “random assignment” was actually random!

Exercise 1

Check for balance in terms of the average values of applicant gender (`female`), computer skills (`computerskills`), and years of experience (`yearsexp`) across the two arms of the experiment (i.e. by `black`). Calculate both the differences in means across treatment arms *and* test for statistical significance of these differences. Do gender and computer skills look balanced across race groups?

Exercise 2

Do a similar tabulation for education (`education`). Education is a categorical variable coded as follows:

- 0: Education not reported
- 1: High school dropout
- 2: High school graduate
- 3: Some college
- 4: College graduate or higher

Because these are categorical, you shouldn’t just calculate and compare means—you should compare share or count of observations with each value (e.g., a chi-squared contingency table). You may also find the `pd.crosstab` function useful.

Does education look balanced across racial groups?

Exercise 3

What do you make of the overall results on resume characteristics? Why do we care about whether these variables look similar across the race groups? And if they didn't look similar, would that be a threat to internal or external validity?

..

Estimating Effect of Race

Exercise 4

The variable of interest in the data set is the variable `call`, which indicates a call back for an interview. Perform a two-sample t-test comparing applicants with black sounding names and white sounding names.

Interpret your results—in both percentage terms *and* in terms of percentage points, what is the effect of having a Black-sounding name (as opposed to a White-sounding name) on your resume?

Exercise 5

Now, use a linear probability model (regression!) to estimate the differential likelihood of being called back by applicant race (i.e. the racial discrimination by employers).

Since we have a limited dependent variable, be sure to use heteroskedastic robust standard errors. [https://www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.OLSResults.get_robustcov_results.html]

Personally, I prefer the HC3 implementation, as it tends to do better with smaller samples than other implementations.

Interpret these results—what is the *effect* of having a Black-sounding name (as opposed to a White-sounding name) on your resume in terms of the likelihood you'll be called back?

Exercise 6

Now let's see if we can improve our estimates by adding in other variables as controls. Add in `education`, `yearsexp`, `female`, and `computerskills`—be sure to treat `education` as a categorical variable!

Estimating Heterogeneous Effects

Exercise 7

What we've been estimating up until this point are the *average* effects. Now let's look for evidence of *heterogeneous treatment effects*—effects that are different for different types of people in our data.

Is there more or less racial discrimination among applicants who do *not* have a college degree? What is the difference in both percentage terms and in percentage points? Is the difference statistically significant?

Please still include `education`, `yearsexp`, `female`, and `computerskills` as controls.

(Hint: use an interaction term)

Exercise 8

Now let's compare men and women—is the penalty for having a Black-sounding

name greater for Black men or Black women?

Again, please still include education, yearsexp, female, and computerskills as controls.

Exercise 9

Calculate and/or lookup the following online:

- What is the share of applicants in our dataset with college degrees?
- What share of Black adult Americans have college degrees (i.e. have completed a bachelors degree)?

Exercise 10

Bearing in mind your answers to Exercise 7 and to Exercise 9, how do you think the Average Treatment Effect you estimated in Exercise 6 might generalize to the experience of the average Black American (i.e., how do you think the ATE for the average Black American would compare to the ATE estimated from this experiment)?

Exercise 11

What does your answer to Exercise 10 imply about the study's *internal* validity?

Exercise 12

What does your answer to Exercise 10 imply about the study's *external* validity?

What Did We Just Measure?

It's worth pausing for a moment to think about exactly what we've measured in this experiment. Was it the effect of race on hiring? Or the difference in the experience of the average White job applicant from the average Black job applicant?

Well... no. What we have measured in this experiment is **just** the effect of having a Black-sounding name (as opposed to a White-sounding name) on your resume on the likelihood of getting a followup call from someone hiring in Boston or Chicago given identical resumes. In that sense, what we've measured is a small *piece* of the difference in the experience of Black and White Americans when seeking employment. As anyone looking for a job knows, getting a call-back is obviously a crucial step in getting a job, so this difference—even if it's just one part of the overall difference—is remarkable.