

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. Also, please note I subscribe to the logic of [Kwame Anthony Appiah](#) and chose to capitalize both the B in Black and the W in White).

You can get access to original article [here](#).

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](#) 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.

Gradescope Autograding

Please follow [all standard guidance](#) for submitting this assignment to the Gradescope autograder, including storing your solutions in a dictionary called `results` and ensuring your notebook runs from the start to completion without any errors.

For this assignment, please name your file `exercise_resume_experiment.ipynb` before uploading.

You can check that you have answers for all questions in your `results` dictionary with this code:

```

assert set(results.keys()) == {
    "ex2_pvalue_computerskills",
    "ex2_pvalue_female",
    "ex2_pvalue_yearsexp",
    "ex3_pvalue_education",
    "ex4_validity",
    "ex5_pvalue",
    "ex5_white_advantage_percent",
    "ex5_white_advantage_percentage_points",
    "ex6_black_pvalue",
    "ex8_black_college",
    "ex8_black_nocollege",
    "ex8_college_heterogeneity",
    "ex9_gender_and_discrimination",
    "ex10_experiment_v_us",
}

```

Submission Limits

Please remember that you are **only allowed THREE submissions to the autograder**. Your last submission (if you submit 3 or fewer times), or your third submission (if you submit more than 3 times) will determine your grade. Submissions that error out will **not** count against this total.

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

Download the data set from this experiment (`resume_experiment.dta`) from [github](#). To aid the autograder, please load the data directly from a URL.

```
In [3]: import pandas as pd
import numpy as np

db_resume = pd.read_stata('https://github.com/nickeubank/MIDS_Data/'
                          'raw/refs/heads/master/resume_experiment/'
                          'resume_experiment.dta')

print(db_resume.columns)
db_resume.sample(2)
```

Index(['education', 'ofjobs', 'yearsexp', 'computerskills', 'call', 'female',
 'black'],
 dtype='object')

```
Out[3]:
```

	education	ofjobs	yearsexp	computerskills	call	female	black
2600	4	6	9	1	0.0	1.0	0.0
4436	4	5	7	1	0.0	1.0	1.0

Exercise 2

- `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.

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. Does gender, computer skills, and yearsexp look balanced across race groups in terms of both statistical significance and magnitude of difference?

Store the p-values associated with your t-test of these variables in `ex2_pvalue_female` , `ex2_pvalue_computerskills` , and `ex2_pvalue_yearsexp` . **Round your values to 2 decimal places.**

Exercise 3

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?

Store the p-value from your chi squared test in results under the key `ex3_pvalue_education` . **Please round to 2 decimal places.**

Exercise 4

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?

Answer in markdown, then also store your answer to the question of whether imbalances are a threat to internal or external validity in `"ex4_validity"` as the string `"internal"` or `"external"` .

Estimating Effect of Race

Exercise 5

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

Store how much more likely a White applicant is to receive a call back than a Black respondent in percentage and percentage points in

`"ex5_white_advantage_percent"` and

`"ex5_white_advantage_percentage_points"` . Please scale percentages so 1 is 1% and percentage points so a value of `1` corresponds to 1 percentage point. **Please round these answers to 2 decimal places.**

Store the p-value of the difference in `"ex5_pvalue"` **Please round your p-value to 5 decimal places.**

Exercise 6

Now, use a linear probability model (a linear regression with a 0/1 dependent variable!) to estimate the differential likelihood of being called back by applicant race (i.e. the racial discrimination by employers). Please use [statsmodels](#).

Since we have a limited dependent variable, be sure to use [heteroskedastic robust standard errors](#). 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?

How does this compare to the estimate you got above in exercise 5?

Store the p-value associated with `black` in `"ex6_black_pvalue"` . **Please round your pvalue to 5 decimal places.**

Exercise 7

Even when doing a randomized experiment, adding control variables to your regression *can* improve the statistical efficiency of your estimates of the treatment effect (the upside is the potential to explain residual variation; the downside is more parameters to be estimated). Adding controls can be particularly useful when randomization left some imbalances in covariates (which you may have seen above).

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 8

As you may recall from some past readings (such as this one on the [migraine medication Aimovig](#)), our focus on estimating *Average Treatment Effects* runs the risk of papering over variation in how individuals respond. In the case of Aimovig, for example, nearly no patients actually experienced the Average Treatment Effect of the medication; around

half of patients experienced no benefit, while the other half experienced a benefit of about twice the average treatment effect.

So far in this analysis we've been focusing on the *average* effect of having a Black-sounding name (as compared to a White-sounding name). But we can actually use our regression framework to look for evidence of *heterogeneous treatment effects*—effects that are different for different types of people in our data. We accomplish this by *interacting* a variable we think may be related to experiencing a differential treatment effect with our treatment variable. For example, if we think that applicants with Black-sounding names who have a college degree are likely to experience less discrimination, we can interact `black` with an indicator for having a college degree. If having a college degree reduces discrimination, we could expect the interaction term to be positive.

Is there more or less racial discrimination (the absolute magnitude difference in call back rates between Black and White applicants) among applicants who have a college degree? Store your answer as the string `"more discrimination"` or `"less discrimination"` under the key `"ex8_college_heterogeneity"`.

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

Note: it's relatively safe to assume that someone hiring employees who sees a resume that does *not* report education levels will assume the applicant does not have a college degree. So treat "No education reported" as "not having a college degree."

In percentage points, what is the difference in call back rates:

- between White applicants without a college degree and Black applicants without a college degree (`ex8_black_nocollege`).
- between White applicants with a college degree and Black applicants with a college degree (`ex8_black_college`).

Use negative values to denote a lower probability for Black applicants to get a call back.

Scale so a value of 1 is a one percentage point difference. Please round your answer to 2 percentage points.

Focus on the coefficient values, even if the significance is low.

Exercise 9

Now let's compare men and women—is the penalty for having a Black-sounding name greater for Black men or Black women? Store your answer as `"greater discrimination for men"` or `"greater discrimination for women"` in `"ex9_gender_and_discrimination"`.

Focus on the coefficient values, even if the significance is low.

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

Exercise 10

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)?

Is the share of Black applicants with college degrees in this data `"greater"`, or `"less"` than in the US? Store your answer as one of those strings in `"ex10_experiment_v_us"`

Exercise 11

Bearing in mind your answers to Exercise 8 and to Exercise 10, how do you think the Average Treatment Effect you estimated in Exercises 5 and 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 12

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

Exercise 13

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.

