

Resume Experiment Analysis

Yuanjing Zhu

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import ttest_ind, chi2_contingency
import statsmodels.api as sm
import statsmodels.formula.api as smf
import warnings
warnings.filterwarnings('ignore')
```

Exercise 1: check balance for gender, computer skills and years of experience

```
In [ ]: # Load dataset
df = pd.read_stata('https://github.com/nickeubank/MIDS_Data/blob/'
                  'master/resume_experiment/resume_experiment.dta?raw=true')
df.head()
```

```
Out[ ]:
```

	education	ofjobs	yearsexp	computerskills	call	female	black
0	4	2	6	1	0.0	1.0	0.0
1	3	3	6	1	0.0	1.0	0.0
2	4	1	6	1	0.0	1.0	1.0
3	3	4	6	1	0.0	1.0	1.0
4	3	3	22	1	0.0	1.0	0.0

```
In [ ]: df.groupby('black')['female', 'computerskills', 'yearsexp'].mean()
```

```
Out[ ]:
```

	female	computerskills	yearsexp
black			
0.0	0.763860	0.808624	7.856263
1.0	0.774538	0.832444	7.829569

```
In [ ]: df_black = df.loc[df['black']==1, ['female', 'computerskills', 'yearsexp']]
df_white = df.loc[df['black']==0, ['female', 'computerskills', 'yearsexp']]

# calculate the difference in means
for i in range(3):
    feature = df_black.columns[i]
    mean_diff = df_black.iloc[:,i].mean() - df_white.iloc[:,i].mean()
    print(f'Differences in means for {feature}: {mean_diff:.2f}')
```

```
Differences in means for female: 0.01
Differences in means for computerskills: 0.02
Differences in means for yearsexp: -0.03
```

```
In [ ]: # test for statistic significance
```

```
_, p = ttest_ind(df_black, df_white, axis=0)
for i in range(3):
    print(f'P-value for {df_black.columns[i]}: {p[i]:.2f}')
```

P-value for female: 0.38
P-value for computerskills: 0.03
P-value for yearsexp: 0.85

The differences in means across treatment arms for female, computerskills, and years of experience were small, with a value of 0.01, 0.02, and -0.03 respectively.

The p-value for female was 0.38, indicating that there was no statistically significant difference in the proportion of female applicants between the two race groups. The p-value for computerskills was 0.03, suggesting that there was a statistically significant difference in the average computer skills of applicants between the two race groups. The p-value for yearsexp was 0.85, indicating that there was no statistically significant difference in the average years of experience between the two race groups. Overall, gender looks balanced across race groups while computer skills may be slightly imbalanced.

Exercise 2: check balance for education

```
In [ ]: df['education_cat'] = df['education'].map({0: 'Education not reported', \
                                                1: 'High school dropout', \
                                                2: 'High school graduate', \
                                                3: 'Some college', \
                                                4: 'College graduate or higher'})
```

```
In [ ]: # chi-square test
cross_tab = pd.crosstab(df['education_cat'], df['black'])
_, p, _, _ = chi2_contingency(cross_tab)
print(f"P-value for chi-square test: {p:.2f}")
```

P-value for chi-square test: 0.49

Based on the result of chi-squared test, education is balanced across the black and white groups (p-value = 0.49).

Exercise 3

Based on t test and chi-square test, there is no significant statistical difference between black and non-black resumes in terms of gender, years of experience, and education. However, there is slight imbalance in computer skills between the two groups.

It is important to check the balance of the covariates in the experiment. If the covariates are not balanced, the results of the experiment may not be biased. In this case, the imbalance in computer skills between the two groups may affect the validity of the experiment. For instance, the slightly higher computer skills of black applicants may lead to a higher probability of getting a callback, which can confound the effect of race on callback rate. Therefore, the imbalance of covariates could be a threat to internal validity because it could lead to ineffective randomization process as well as biased estimates of the treatment effect. On the other hand, if the two groups are not balanced, it may be a threat to external validity since the results of the study might not generalize to other populations with different distributions of these characteristics.

Exercise 4: effect of black-sounding name on callback rate

```
In [ ]: black_call = df.loc[df['black']==1, 'call']
white_call = df.loc[df['black']==0, 'call']

# perform ttest
_, p = ttest_ind(black_call, white_call)
# print results
print(f"P-value for t-test: {p:.2f}")
if p < 0.05:
    print('Reject null hypothesis, there is a difference in call back rate')
else:
    print('Fail to reject null hypothesis, there is no difference in call back rate')
```

P-value for t-test: 0.00

Reject null hypothesis, there is a difference in call back rate

```
In [ ]: # percentage and percentage points
black_call_rate = black_call.mean()*100
white_call_rate = white_call.mean()*100
percentage_diff = (black_call_rate-white_call_rate)
percentage_change = percentage_diff / white_call_rate * 100
print(f"Percentage of black resumes that received a call: \
      {black_call_rate:.2f}%")
print(f"Percentage of white resumes that received a call: \
      {white_call_rate:.2f}%")
print(f"Percentage points difference in resumes that received a call: \
      {percentage_diff:.2f}%")
print(f"Percentage terms of resumes that received a call: \
      {percentage_change:.2f}%")
```

Percentage of black resumes that received a call: 6.45%

Percentage of white resumes that received a call: 9.65%

Percentage points difference in resumes that received a call: -3.20%

Percentage terms of resumes that received a call: -33.19%

The two-sample t-test suggests that there is statistically significant difference in the callback rate between black and non-black applicants ($p\text{-value} < 0.05$). The callback rate of black applicants is 6.45% while the callback rate of non-black applicants is 9.65%. Therefore, having a Black-sounding name (as opposed to a White-sounding name) on a resume reduced the probability of receiving a call for an interview by 3.20 percentage points. This means that the percentage of black resumes that received a call was 33.19% lower than the percentage of white resumes that received a call.

Exercise 5: use a linear probability model to estimate the differential likelihood of being called back by applicant race

```
In [ ]: model_5 = smf.ols("call ~ black", df).fit()
model_5.get_robustcov_results(cov_type="HC3").summary()
```

Out[]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.003
Model:	OLS	Adj. R-squared:	0.003
Method:	Least Squares	F-statistic:	16.92
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	3.96e-05
Time:	21:09:03	Log-Likelihood:	-562.24
No. Observations:	4870	AIC:	1128.
Df Residuals:	4868	BIC:	1141.
Df Model:	1		
Covariance Type:	HC3		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0965	0.006	16.121	0.000	0.085	0.108
black	-0.0320	0.008	-4.114	0.000	-0.047	-0.017
Omnibus:	2969.205	Durbin-Watson:	1.440			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18927.068			
Skew:	3.068	Prob(JB):	0.00			
Kurtosis:	10.458	Cond. No.	2.62			

Notes:

[1] Standard Errors are heteroscedasticity robust (HC3)

```
In [ ]: print(f"Coefficient of black is: {model_5.params['black']:.2f}")
```

Coefficient of black is: -0.03

The linear regression model shows that having a Black-sounding name (as opposed to a White-sounding name) is associated with a statistically significant decrease in the likelihood of being called back. Holding all other variables constant, a Black applicant is about 3% less likely to receive a call back compared to a White applicant.

Exercise 6: Add in education, yearsexp, female, and computerskills

```
In [ ]: model_6=smf.ols("call~black + C(education) + yearsexp "+  
                        "+ female + computerskills",df).fit()  
model_6.get_robustcov_results(cov_type="HC3").summary()
```

Out[]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.006
Method:	Least Squares	F-statistic:	4.350
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	3.04e-05
Time:	21:09:03	Log-Likelihood:	-551.02
No. Observations:	4870	AIC:	1120.
Df Residuals:	4861	BIC:	1178.
Df Model:	8		
Covariance Type:	HC3		
	coef	std err	t P> t [0.025 0.975]
Intercept	0.0821	0.040	2.053 0.040 0.004 0.160
C(education)[T.1]	-0.0017	0.057	-0.030 0.976 -0.113 0.110
C(education)[T.2]	-8.953e-05	0.042	-0.002 0.998 -0.082 0.082
C(education)[T.3]	-0.0025	0.039	-0.065 0.948 -0.079 0.074
C(education)[T.4]	-0.0047	0.038	-0.124 0.901 -0.080 0.070
black	-0.0316	0.008	-4.076 0.000 -0.047 -0.016
yearsexp	0.0032	0.001	3.665 0.000 0.001 0.005
female	0.0112	0.010	1.165 0.244 -0.008 0.030
computerskills	-0.0186	0.011	-1.616 0.106 -0.041 0.004
Omnibus:	2950.646	Durbin-Watson:	1.448
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18631.250
Skew:	3.047	Prob(JB):	0.00
Kurtosis:	10.395	Cond. No.	225.

Notes:

[1] Standard Errors are heteroscedasticity robust (HC3)

Compared with the last mode, R^2 increased from 0.003 to 0.008 after adding other variables, indicating the model is improved. Among the 5 variables, only black and years of experience are statistically significant. Holding all other variables constant, a Black applicant is about 3% less likely to receive a call back compared to a White applicant, which is similar to the result in exercise 5. On the other hand, holding all other variables constant, a one year increase in years of experience is associated with a 0.3% increase in the likelihood of being called back.

Exercise 7: racial discrimination among applicants who do not have a college degree

```
In [ ]: df['college_degree'] = np.where (df['education'] >= 4, 1, 0)
model = smf.ols("call ~ C(black) + C(college_degree) + C(black):C(college_degree)"
               "+ C(education) + yearsexp + C(female) + computerskills", df).fit()
model.get_robustcov_results(cov_type="HC3").summary()
```

Out[]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.006
Method:	Least Squares	F-statistic:	3.952
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	4.93e-05
Time:	21:09:03	Log-Likelihood:	-550.76
No. Observations:	4870	AIC:	1122.
Df Residuals:	4860	BIC:	1186.
Df Model:	9		
Covariance Type:	HC3		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0875	0.040	2.176	0.030	0.009	0.166
C(black)[T.1.0]	-0.0405	0.015	-2.736	0.006	-0.070	-0.011
C(college_degree)[T.1]	-0.0060	0.019	-0.307	0.759	-0.044	0.032
C(education)[T.1]	-0.0023	0.057	-0.040	0.968	-0.114	0.110
C(education)[T.2]	-0.0012	0.042	-0.030	0.976	-0.083	0.081
C(education)[T.3]	-0.0036	0.039	-0.092	0.927	-0.080	0.073
C(education)[T.4]	-0.0060	0.019	-0.307	0.759	-0.044	0.032
C(female)[T.1.0]	0.0112	0.010	1.157	0.247	-0.008	0.030
C(black)[T.1.0]:C(college_degree)[T.1]	0.0123	0.017	0.710	0.478	-0.022	0.046
yearsexp	0.0032	0.001	3.672	0.000	0.001	0.005
computerskills	-0.0186	0.011	-1.618	0.106	-0.041	0.004

Omnibus:	2950.182	Durbin-Watson:	1.448
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18623.859
Skew:	3.046	Prob(JB):	0.00
Kurtosis:	10.393	Cond. No.	4.97e+17

Notes:

[1] Standard Errors are heteroscedasticity robust (HC3)

[2] The smallest eigenvalue is 1.76e-30. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

The coefficient of black, college degree, and the interaction term between black and college degree is -0.0405, -0.006, and 0.0123 respectively.

For those who do not have college degree, having a black-sounding name will have 4.5% lower callback rate than having a white-sounding name.

For those who have college degree, having a black-sounding name will have 2.82% lower callback rate than having a white-sounding name ($-0.0405 + 0.0123 = -0.0282$)

Therefore, **there are more racial discrimination among applicants who do not have a college degree.**

The difference in percent point of having a black name on college degree is 1.68% ($0.045 - 0.0282 = 0.0168$). The percentage term is 37% ($0.0168/0.045 = 0.037$).

However, the p-value of the interaction term suggests that this difference is not statistically significant, indicating that we cannot rule out the possibility that this difference occurred by chance.

Note: the feature "college_degree" is generated from the feature "education". Including both might cause multicollinearity problem.

Exercise 8: is the penalty for having a Black-sounding name greater for Black men or Black women?

```
In [ ]: model = smf.ols("call ~ C(black) + C(female) + C(black):C(female) + "
                    "C(education) + yearsexp + computerskills", df).fit()
model.get_robustcov_results(cov_type="HC3").summary()
```

Out[]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.006
Method:	Least Squares	F-statistic:	3.866
Date:	Mon, 20 Feb 2023	Prob (F-statistic):	6.76e-05
Time:	21:09:03	Log-Likelihood:	-551.00
No. Observations:	4870	AIC:	1122.
Df Residuals:	4860	BIC:	1187.
Df Model:	9		
Covariance Type:	HC3		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0807	0.040	1.996	0.046	0.001	0.160
C(black)[T.1.0]	-0.0287	0.016	-1.840	0.066	-0.059	0.002
C(female)[T.1.0]	0.0131	0.014	0.919	0.358	-0.015	0.041
C(education)[T.1]	-0.0021	0.057	-0.037	0.971	-0.114	0.110
C(education)[T.2]	-0.0001	0.042	-0.003	0.998	-0.082	0.082
C(education)[T.3]	-0.0026	0.039	-0.066	0.947	-0.079	0.074
C(education)[T.4]	-0.0048	0.038	-0.125	0.900	-0.080	0.070
C(black)[T.1.0]:C(female)[T.1.0]	-0.0038	0.018	-0.213	0.831	-0.039	0.031
yearsexp	0.0032	0.001	3.668	0.000	0.001	0.005
computerskills	-0.0186	0.011	-1.618	0.106	-0.041	0.004

Omnibus:	2950.616	Durbin-Watson:	1.448
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18630.964
Skew:	3.047	Prob(JB):	0.00
Kurtosis:	10.395	Cond. No.	226.

Notes:

[1] Standard Errors are heteroscedasticity robust (HC3)

The coefficient of black, female, and the interaction term between black and female is -0.0287, 0.0131, and -0.0038 respectively.

For male, having a black-sounding name will have 2.87% lower callback rate than having a white-sounding name.

For female, having a black-sounding name will have 3.25% lower callback rate than having a white-sounding name ($-0.0287 + -0.0038 = -0.0325$)

Therefore, **having a black-sounding name has more penalty among female applicants.**

However, the p-value of the interaction term suggests that this difference is not statistically significant, indicating that we cannot rule out the possibility that this difference occurred by chance.

Exercise 9

```
In [ ]: college_share = df['college_degree'].mean()
print(f"Percentage of college degrees in our dataset: \
      {college_share*100:.2f}%")

black = df[df['black'] == 1]
black_college_share = (black['college_degree']).mean()
print(f"Percentage of college degrees in our black dataset: \
      {black_college_share*100:.2f}%")
```

```
Percentage of college degrees in our dataset:      71.95%
Percentage of college degrees in our black dataset:  72.28%
```

According to [US Census Bureau](#), the percentage of adults age 25 and older with a bachelor's degree or higher for the Black population is 28.1% in 2021.

Exercise 10

Given that the percentage of college degrees in the dataset is 71.95%, and the percentage of college degrees in our black dataset is 72.28%, it is likely that the experiment is not representative of the general population. The share of Black adult Americans who have college degrees is 28.1%, which is significantly lower than the percentage of college degrees in the dataset. This means that the ATE estimated from this experiment may not necessarily generalize to the experience of the average Black American who does not have a college degree. Since the average Black American is less likely to have a college degree than the applicants in this experiment, the ATE estimated from this experiment may overestimate the effect of having a college degree on the outcome of interest for the average Black American.

Exercise 11

Internal validity is about whether a study has accurately measured a causal effect in the context being studied. It is not related to how conclusions from the study can be generalized to other contexts. Therefore, the internal validity of this study is not affected by the fact that the study is not representative of the general population.

However, the sample imbalance of computer skills may affect the internal validity of the study. If the imbalance in computer skills between the two groups is not random, it may lead to ineffective randomization process as well as biased estimates of the treatment effect.

Exercise 12

External validity refers to the extent to which the findings from a study can be generalized to other populations, settings, and times. The answer to Exercise 10 suggests that the study's external validity may be limited. The fact that the percentage of college degrees in the dataset is much higher than the percentage of average Black Americans with college degrees suggests that the study sample may not be representative of the overall population, and the ATE estimated from this experiment may not necessarily generalize to the experience of the average Black American who does not have a college degree. Therefore, the external validity of this study is limited.