Resume Experiment Analysis

Assignment descriptions:

https://www.unifyingdatascience.org/html/exercises/exercise_resume.html

Checking for Balance

Exercise 1

Check for balance in terms 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 across treatment arms and test for statistical significance of these differences. Do gender and computer skills look balanced across race groups? (1 point)

```
In [1]:  import pandas as pd
         import numpy as np
         from scipy import stats
         from sklearn.linear_model import LinearRegression
```

```
In [2]:  df = pd.read_stata("resume_experiment.dta")
         df.head()
```

Out[2]:

	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 [3]:  treat = df[df['black'] == 1]
         ctrl = df[df['black'] == 0]
```

```
In [4]:  meanT = treat[['female', 'computerskills', 'yearsexp']].mean()
         meanC = ctrl[['female', 'computerskills', 'yearsexp']].mean()
         print('Mean difference in terms of features across groups:')
         print(meanT - meanC)
```

```
Mean difference in terms of features across groups:
female             0.010678
computerskills     0.023819
yearsexp          -0.026694
dtype: float64
```

```
In [5]:  print('In terms of gender:')
         print(stats.ttest_ind(treat['female'], ctrl['female']), '\n')

         print('In terms of computer skills:')
         print(stats.ttest_ind(treat['computerskills'], ctrl['computerskills']), '\n')

         print('In terms of years of experience:')
         print(stats.ttest_ind(treat['yearsexp'], ctrl['yearsexp']))
```

```
In terms of gender:
Ttest_indResult(statistic=0.8841321018026016, pvalue=0.37666856909823254)
```

```
In terms of computer skills:
Ttest_indResult(statistic=2.1664271042751966, pvalue=0.030326933955391936)

In terms of years of experience:
Ttest_indResult(statistic=-0.18461970685747395, pvalue=0.8535350182481283)
```

> In particular, the feature gender can be regarded as balanced across race groups
> due to the large p-value of the t-test. However, the feature of computer skills may
> look balanced if we set the significance level to be 0.01 but may also become
> unbalanced if the threshold is more than 0.05.

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 of observations with each value separately using a ttest, or do a chi-squared test
(technically chi-squared is the correct test, but I'm ok with either).

Does education and the number of previous jobs look balanced across racial groups? (2 points)

In [6]:
```python
# education
obs1 = np.array([treat['education'].value_counts(), ctrl['education']. \
                value_counts()])
chi2, p, dof, expected = stats.chi2_contingency(obs1)
print(f"p value is {p}.")
```

p value is 0.4917640058792272.

In [7]:
```python
# number of previous jobs
obs2 = np.array([treat['ofjobs'].value_counts(), ctrl['ofjobs']. \
                value_counts()])
stat, p, dof, expected = stats.chi2_contingency(obs2)
print(f"p value is {p}.")
```

p value is 0.7406654986208298.

> Given the large p-values (> 0.1) from both Chi-squared tests, it is safe to conclude
> that education and the number of previous jobs are balanced across 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? (1 point)

> Overall, the features of this resume experiment are mostly balanced except for
> computer skills, which can also be considered balance if we take 0.01 significance
> level. It is important to ensure the underlying features are balanced so that the
> experiment is free from selection bias. Thereby, our causal inference is valid.

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.

```
In [8]:  print(stats.ttest_ind(treat['call'], ctrl['call']))
```

```
Ttest_indResult(statistic=-4.114705290861751, pvalue=3.940802103128886e-05)
```

> The extremely low p-value indicates that the chances for Black-named applicants to receive callbacks for an interview are different from White-named applicants in this fictious experiment.

Exercise 5

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

```
In [9]:  import statsmodels.api as sm
         import statsmodels.formula.api as smf
         smf.ols("call ~ black", df).fit().summary()
```

Out[9]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.003
Model:	OLS	Adj. R-squared:	0.003
Method:	Least Squares	F-statistic:	16.93
Date:	Tue, 09 Feb 2021	Prob (F-statistic):	3.94e-05
Time:	17:36:39	Log-Likelihood:	-562.24
No. Observations:	4870	AIC:	1128.
Df Residuals:	4868	BIC:	1141.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0965	0.006	17.532	0.000	0.086	0.107
black	-0.0320	0.008	-4.115	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 assume that the covariance matrix of the errors is correctly specified.

> According to the regression model, Black-named fictitious applicants are expected to receive 0.03 less calls than White-named applicants. In this fictitious experiment, there is racial discrimination by employers.

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!

In [10]:
```
# treat education as a categorical variable
df['education'] = df['education'].astype('category')
```

In [11]:
```
smf.ols("call ~ black + education + yearsexp + female + computerskills", df) \
    .fit().summary()
```

Out[11]:

OLS Regression Results

Dep. Variable:	call	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.006
Method:	Least Squares	F-statistic:	4.931
Date:	Tue, 09 Feb 2021	Prob (F-statistic):	4.30e-06
Time:	17:36:39	Log-Likelihood:	-551.02
No. Observations:	4870	AIC:	1120.
Df Residuals:	4861	BIC:	1178.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0821	0.042	1.957	0.050	-0.000	0.164
education[T.1]	-0.0017	0.059	-0.029	0.977	-0.117	0.113
education[T.2]	-8.953e-05	0.044	-0.002	0.998	-0.086	0.085
education[T.3]	-0.0025	0.041	-0.061	0.951	-0.083	0.078
education[T.4]	-0.0047	0.040	-0.117	0.907	-0.084	0.074
black	-0.0316	0.008	-4.064	0.000	-0.047	-0.016
yearsexp	0.0032	0.001	4.067	0.000	0.002	0.005
female	0.0112	0.010	1.153	0.249	-0.008	0.030
computerskills	-0.0186	0.011	-1.743	0.081	-0.039	0.002

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 assume that the covariance matrix of the errors is correctly specified.

> It is found that the standard error of the black feature remains as 0.008 with more covariates added. Also, the 95% confidence interval of estimated black variable remains approximately the same. Therefore, the estimation in terms of racial effects is not improved. However, the adjusted R-squared increases from 0.003 to 0.006, which suggests improved fitting generally.

Estimating Heterogeneous Effects

Exercise 7

These effects are the average effects. Now let's look for heterogeneous treatment effects.

Look only at candidates with high educations. Is there more or less racial discrimination among these highly educated candidates?

> Let's assume applicants with college degrees are candidates with high educations.

```
In [12]:  high = df[(df['education'] == 3) | (df['education'] == 4)]
          high.head()
```

Out[12]:

	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 [13]:  smf.ols("call ~ black + education + yearsexp + female + computerskills", high) \
              .fit().summary()
```

Out[13]:

OLS Regression Results

Dep. Variable:	call	**R-squared:**	0.007
Model:	OLS	**Adj. R-squared:**	0.006
Method:	Least Squares	**F-statistic:**	6.795
Date:	Tue, 09 Feb 2021	**Prob (F-statistic):**	2.54e-06
Time:	17:36:39	**Log-Likelihood:**	-500.06
No. Observations:	4510	**AIC:**	1012.
Df Residuals:	4504	**BIC:**	1051.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0504	0.010	5.069	0.000	0.031	0.070
education[T.1]	1.671e-16	3.39e-17	4.925	0.000	1.01e-16	2.34e-16
education[T.2]	1.263e-17	2.8e-18	4.515	0.000	7.15e-18	1.81e-17
education[T.3]	0.0256	0.008	3.082	0.002	0.009	0.042
education[T.4]	0.0248	0.006	4.453	0.000	0.014	0.036
black	-0.0329	0.008	-4.079	0.000	-0.049	-0.017
yearsexp	0.0028	0.001	3.475	0.001	0.001	0.004
female	0.0164	0.010	1.611	0.107	-0.004	0.036
computerskills	-0.0163	0.011	-1.449	0.147	-0.038	0.006

Omnibus:	2747.273	Durbin-Watson:	1.444
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17523.786
Skew:	3.061	Prob(JB):	0.00
Kurtosis:	10.469	Cond. No.	5.43e+17

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.36e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

> Unfortunately, racial discrimination is more severe among candidates with high education. According to the regression results, highly educated black applicants are expected to receive 0.0329 less callbacks. This means the opportunities are 4% less if black people acquire high education in this study.

Exercise 8

Now let's compare men and women – is discrimination greater for Black men or Black women?

> Given our regression results in Exercise 6, the estimated coefficient for females is positive, 0.0112, which implies women have higher chances of callbacks regardless of racial groups. Therefore, the racial discrimination for Black women and men isn't greater in terms of genders.

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)?

```
In [14]: share = round(high.shape[0]/df.shape[0], 4)
```

```
print(f'The share of applicants in our dataset with college degrees is {share}.')
```

The share of applicants in our dataset with college degrees is 0.9261.

In [15]:
```
treat_high = treat[(treat['education'] == 3) | (treat['education'] == 4)]
share_treat = round(treat_high.shape[0]/treat.shape[0], 4)
print(f'The share of Black adult Americans have college degrees is {share_treat}.')
```

The share of Black adult Americans have college degrees is 0.9253.

Exercise 10

What are the implications of your answers to Exercise 7 and to Exercise 9 to how you interpret the Average Treatment Effect you estimated in Exercise 6?

> The average treatment effect of Black races on callbacks for interviews is not reduced with better educational backgrounds. It can be expected since the majority of fictious candidates in this studies have received high education, which is over 90%. And the proportion of candidates with college degrees is approximately the same within Black community and the whole population in this fictious experiment. Therefore, the estimated treatment effect of Black races looks almost the same no matter the candidates are highly educated or not.