Statistical Inferences

Table of Contents

1. Introduction **2**

2. Dataset **2**

3. Analysis **4**

3.1 Hypothesis 4

3.2 Data and Statistics: 4

3.3 Confidence interval calculation 7

3.4 Statistical significance 8

3.5 Practical significance 8

3.6 Power of the test 8

4. Results and Conclusion **9**

5. Reflection **12**

6. References **11**

7. Appendix **12**

1. Introduction

In this report, we will discuss the difference in income between Black Americans and White Americans in the mid of 1970s. The research's question is: did Black and White Americans have the same income back in 1978, given that members of both groups are degree holders and received the same training program? The report will provide insights into the income disparity at that time. The report will use the data collected in a review article to conduct a difference of means test and compute the test's statistical power and significance power. Thus, we will determine if there is significant evidence to support the hypothesis that black Americans received less income than white Americans because of racial discrimination at that time.

2. Dataset

The data was used from a review article published in the American Economic Review by Robert Lalonde (LaLonde, 1986) in his experiment to see the impact of training programs on trainee earnings. The study asked for the trainee's age, race, years of education, marital status, degree/non-degree holder, and earnings both before and after conducting the training in 1975 and 1978, respectively. The study divided the sample into two groups: the treatment and the control group. The treatment group underwent training to improve their technical skills to be more qualified for well-paid jobs, while the control group did not. However, the data set did not contain the population number of black or white people in the United States. Thus, a report from the United States Census was used to determine the number of white and black people who held degrees in the United States by 1974. The report indicated that the number of black people who earned a degree was 3,105,000, while the number of white people who earned a degree was 22,141,000 (Bureau, 1975).

Based on the research question, which aims to discover the income equality of black and white Americans who had the same qualifications in the mid-1970s, we will focus on these variables:

1- Independent variables

All these variables are qualitative nominal variables. They are qualitative because they group data according to specific criteria. They are nominal because their categorization does not have a degree of ranking. For instance, white people are not higher than black on a particular scale, but they are all equal classifications of humans.

a. **Ethnicity:** the sample will be divided into two groups: black and white people, to determine if ethnicity is the reason behind income inequality if it exists.

b. **Degree:** all the sample will be degree holders because we want to ensure that they are equal in education.

c. **Treatment:** all the samples will be part of the treatment group to ensure they all received the training program and are equal in skills qualifications.

2- Dependent variables

a. **Income in 1978:** this variable will show the impact of the training that acquired both groups with the technical skills essential for the labor market. The variable will be quantitative and continuous. It will be quantitative because adding the values of all white and black people's income and dividing them by the sample number would provide a meaningful value which is the mean. Also, it is continuous because it does not have a precision degree, meaning that the income could be an integer (in the case of 0 income only) or a decimal with a different number of digits. For instance, one sample has an income of 4056.493896 while the other has an income of 480.5270081. This variable will be measured in American dollars.

3. Analysis

To address our question about whether white and black people in the United States had the same income, provided that they both have the same qualifications,

3.1 Hypothesis

The null and alternative hypotheses tested in this paper will be:

H_0: The income of degree-holders and trainee white people is equal to that of degree-holders and trainee black people. ($\mu_{White-}\ \mu_{Black} = 0$)

H_A: The income of degree-holder and trainee white people is higher than that of degree-holders and trainee Black people. ($\mu_{White-}\ \mu_{Black} > 0$)

We will use the difference between the means tests. The sample will be divided into two groups based on their black or white ethnicity. The significance level will be higher than the default $\alpha = 0.1$ because the danger of committing a Type II error is more serious.

As committing Type II error means failing to reject the null hypothesis when it is false, failing to reject the null hypothesis means that we admit that white and black were treated equally in the United States in the past, which might affect our historical perception and knowledge of racial discrimination. The test will be one-sided because we suspect that white people earn more than black people. People with 0 income were removed from both groups because we wanted to examine the income of employed people.

3.2 Data and Statistics:

The data was read into Python and was analyzed using multiple libraries like pandas, stats, and NumPy. The table below, beside the graphs of the two groups, will be used to test the conditions of applying the difference between the two means test. The other statistics of the sample are found in Appendix A:

Table1: Population number and the important statistics for the black and white group		
	White People	**Black people**
Population Number	22,141,000	3,105,000
Sample Number	17	44
Mean	5910.17	8794.68
Standard deviation	4404.74	6304.97

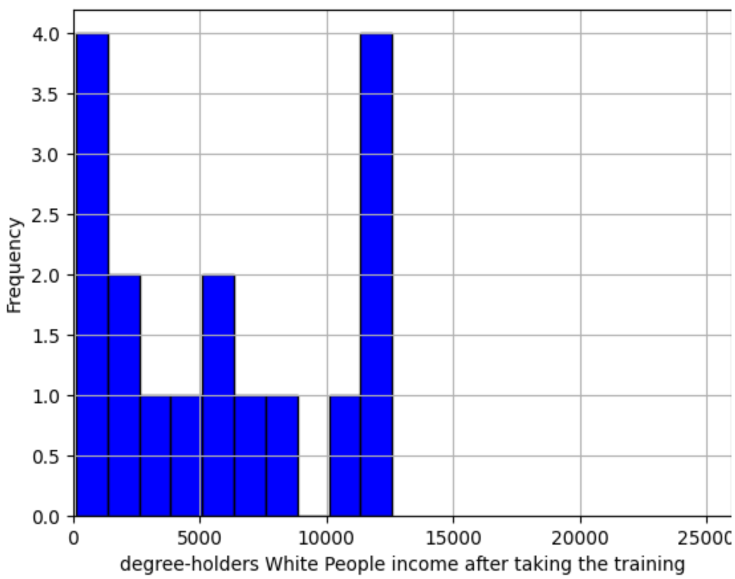

Figure.1
Histogram representing the frequency of white people who hold a degree, income After finishing the training in 1978

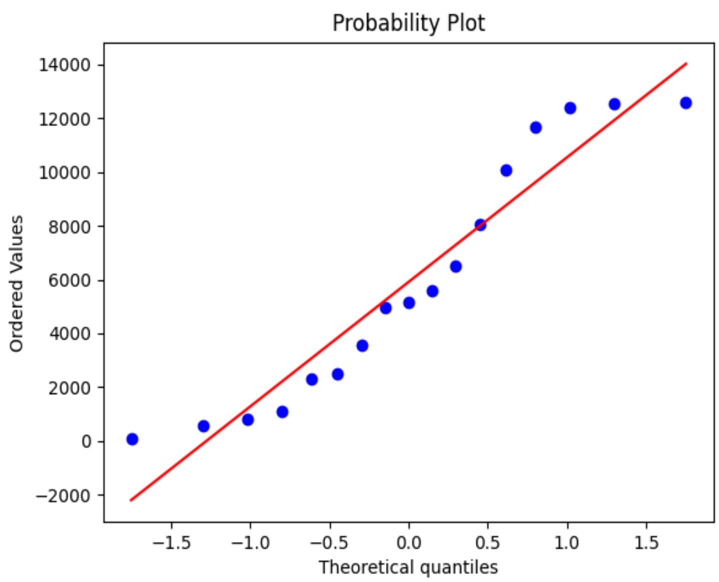

Figure.2
QQ probability plot of the income of the white group.
The graph indicates symmetric lite deviation from the normal distribution

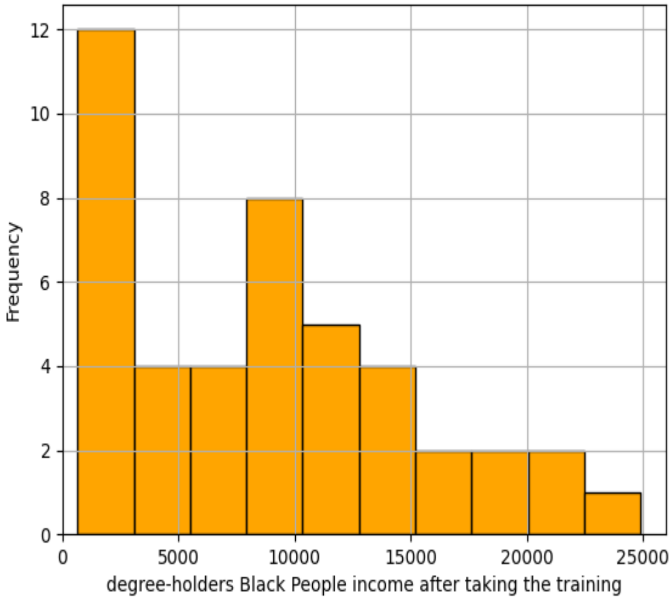

Figure.3
Histogram representing the frequency of Black people who hold a degree, income After finishing the training in 1978.

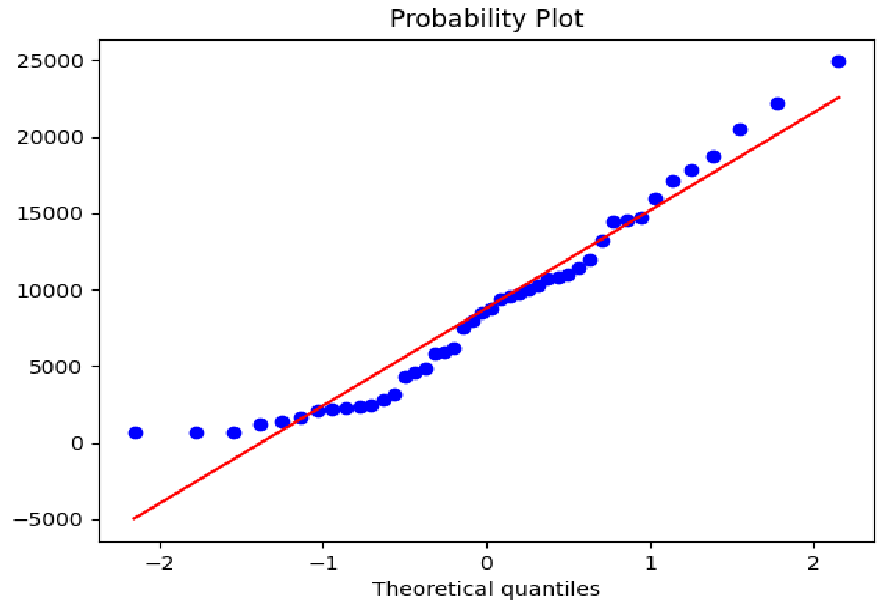

Figure.4
QQ probability plot of the income of the black group.
The graph indicates asymmetric moderate deviation from the normal distribution.

From the statistics and the graphs, we could reach some conclusions:

1- Both groups represent less than 10% of the population, which satisfies the condition of the observation's independence.
2- The black people sample is skewed to the right, but its number is greater than 30, so we can relax the normality condition (Diez et al., 2017).
3- The white people sample is less than 30. However, the sample is symmetric and uniform because of being a bimodal distribution which also means that it satisfies the condition of normality because it is not heavily skewed (McDonald, 2015).

3.3 Confidence interval calculation

Because the sample satisfies the conditions of independence and normality in addition to being randomly chosen according to the study it was picked from (LaLonde, 1986), the relatively small sample size, and the absence of the standard deviation value of the population, we will use the *t*-distribution to calculate the 95 % confidence interval of the difference between the two means of the income of the population of both groups. The full calculations and mathematical equations are found in Appendix B. The confidence interval of the difference between the average income of white and black people rounded to two decimal places [-5913.85, 146.84] American dollars. This means that if we pick multiple samples of black and white people and calculate the difference between the mean of their income and construct confidence intervals for the differences of both populations' mean, the previous interval could be among the 95% intervals that captured the actual population's mean.

3.4 Statistical significance

To calculate the statistical significance of the test, we calculated the p-value. As each group meets the conditions of using the t-distribution as explained above, we can use the t-distribution to find the p-value of the difference between the two means. We will use the same equations in Appendix B to calculate the t-score and the standard error. The calculations in Appendix C show that the t-score is 2.1199, and the standard error is 1429.94. From these values, we can calculate the p-value, which will be equal to 0.03. This p-value represents the probability of observing the difference in the average income between black people and white people observed in 1978, provided that the null hypothesis is true (there was income equality in 1978). Because the p-value is less than the significance level ($\alpha = 0.1$), we have statistically significant evidence to reject the null hypothesis and state that there is a difference between the income of white and black people.

3.5 Practical significance

To further measure the practical significance, we measure the significance level using Hedge's g because it corrects the upward bias presented in small sample sizes like in our case. The equation and full calculations of Hedge's g are in appendix D. Hedge's g resulted in an effect size of - 0.52, which indicates medium negative significance. In other words, it indicates that the average of black and white people's income is not equal, but black people are those with higher incomes.

3.6 Power of the test

We conducted a power test to measure the probability of detecting a difference in the average income of black and white people**.** The calculations are in Appendix E. The null hypothesis was no income difference, and the alternative hypothesis represented the practically significant level of difference from the white group income in 1978, which was calculated by multiplying their average income by the practical significance measured in the previous section. The test showed a power of 99.9%, meaning the probability of committing a Type II error is 0.01%. In other terms, the probability of admitting there was no difference in income in 1978 between both groups while there was an actual difference is almost negligible.

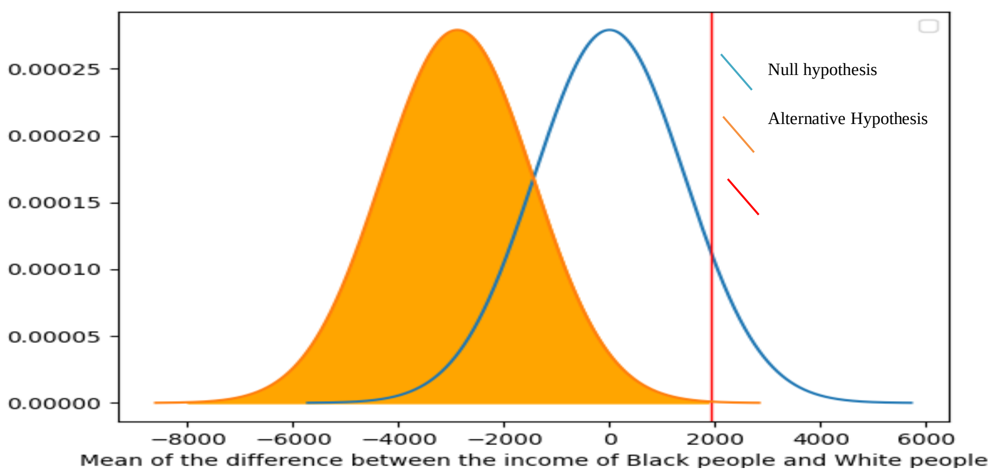

Figure.5
The power of the test. The shaded region represents the power.

4. Results and Conclusion

Based on this study which relied on information from the study of Lalonde and the report of the United States Bureau, we investigated the validity of the claim that there was a racial disparity that caused income inequality between black and white people (Hero & Levy, 2016).

If we assume that the confidence interval of the difference between the mean of the income of both groups was among the 95% of the intervals that held the true population's mean, it did not provide rigorous evidence to refute the null hypothesis or deny the alternative hypothesis because it extended from a negative to a positive range, which could be interpreted as no difference or higher income of either group.

A statistical significance was conducted to further check both hypotheses, resulting in a p-value of 0.03. Thus, we failed to accept the null hypothesis that both groups are equal in income. This fact was also supported by the power test, which showed that we have a 0.01% probability of accepting the null hypothesis while it is false, based on the average income of both groups at that time.

However, we did not have evidence to accept the alternative hypothesis. Based on the calculation of Hedge's g, the effect size was -0.52, which could be interpreted as a moderate significance but in the reverse direction. In other terms, that means that there was moderately-significant evidence to support the hypothesis that there was income disparity, but black people were the group of higher income, not the white, which refutes the alternative hypothesis.

This conclusion is inductive because it used statistical approaches like T-distribution to make general inferences about the population of black and white people in 1978 based on observations collected from a sample of both populations. The conclusion is strong because it was supported using the eliminative approach by providing different forms of evidence, like the statistical significance measured using the p-value and the power of the test, which that the null hypothesis could not be accepted. In contrast, the confidence interval and the practical significance indicated that the alternative hypothesis was inaccurate and that black people earned more than white people in 1978. However, the conclusion is not absolute because another sample might have different statistics, which might affect the statistical and practical significance of the test. The conclusion was reliable because the sample met the conditions of independence, randomness, and normality, which were essential to apply the T-distribution.

Eventually, I advise future researchers to work on larger samples because the small samples might yield inaccurate results that might cause false inferences about the population.

6. References

Bureau, U. C. (1975, July). *The Social and Economic Status of the Black Population in the United States: 1974*. Census.gov. https://www.census.gov/library/publications/1975/demo/p23-054.html

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Journal of the American Statistical Association*, *73*(363), 680. https://doi.org/10.2307/2286629

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2017, August 19). *OpenIntro Statistics*. Www.openintro.org. https://www.openintro.org/book/os/

Hero, R. E., & Levy, M. E. (2016). The Racial Structure of Economic Inequality in the United States: Understanding Change and Continuity in an Era of "Great Divergence." *Social Science Quarterly*, *97*(3), 491–505. https://www.jstor.org/stable/26612333

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, *76*(4), 604–620. https://www.jstor.org/stable/1806062

McDonald, J. H. (2015, July 20). *One-sample t–test - Handbook of Biological Statistics*. Www.biostathandbook.com. http://www.biostathandbook.com/onesamplettest.html

7. Appendix

The full python code and the data can be accessed in the zipped folder submitted as a secondary file.

Appendix A

	White People	Black people
Sample Number	17	44
Mean	5910.17	8794.68
Median	5149.5	8643.95
Mode	N/A	N/A
Standard deviation	4404.74	6304.97
Range	12491.14	24262.244

```
url='https://drive.google.com/file/d/1x333Xl9yNXGFb2v47ZviG2qkjijRrsAj/view?usp=share_link'
url='https://drive.google.com/uc?id=' + url.split('/')[-2]
# read the dataset to make operations on the data and filter it
df = pd.read_csv(url)
# filter the data to include only people who have a degree and received the training
treatment= df.loc[df["treat"]==1]
degree_holder= treatment.loc[treatment["nodegree"]==0]
# Break the dataset into two groups, black group and white group respectively
black_people_group=degree_holder.loc[degree_holder["black"]==1]
white_people_group= degree_holder.loc[degree_holder["black"]==0]
# eliminate people who did not receive income in 1978
white_people_group_income= white_people_group.loc[white_people_group["re78"]>0]
black_people_group_income= black_people_group.loc[ black_people_group["re78"]>0]
# remove the outlier from the dataset
Black_people_group_income_nooutliers= black_people_group_income.loc[ black_people_group_income["re78"] < 30000]
# calculate the sample number of both groups
white_people_group_income=white_people_group_income["re78"]
Black_people_group_income_nooutliers=Black_people_group_income_nooutliers["re78"]
n_white=len(white_people_group_income)
n_black=len(Black_people_group_income_nooutliers)
print("the sample number of the black people group is" +"" + str(n_black))
print("the sample number of the white people group is" + ""+ str(n_white))
# calculate the mean, median, mode, range and standard deviation of the white group
white_mean=np.mean(white_people_group_income)
std_white= np.std(white_people_group_income)
median_white=np.median(white_people_group_income)
range_white= max(white_people_group_income)- min(white_people_group_income)
print("the mean of white people income"+str(white_mean))
print("the Standard deviation of white people income"+str(std_white))
print("the median of white people income"+str(median_white))
print("the range of white people income"+str(range_white))
```

```
black_mean=np.mean(Black_people_group_income_nooutliers)
std_black= np.std(Black_people_group_income_nooutliers)
median_black=np.median(Black_people_group_income_nooutliers)
range_black= max(Black_people_group_income_nooutliers)- min(Black_people_group_income_nooutliers)
print("the mean of black people income"+str(black_mean))
print("the Standard deviation of black people income"+str(std_black))
print("the median of black people income"+str(median_black))
print("the range of black people income"+str(range_black))
# plot histograms for both groups to present the frequency of the income

white_people_group_income.hist(color="b",bins=10,edgecolor="black")
plt.xlabel("degree-holders White People income after taking the training")
plt.ylabel("Frequency")
# xlim was used to unify the x-axis of both groups
# unifying x-axis ease the observation of differences between both graphs
plt.xlim([0,26000])
plt.show()

Black_people_group_income_nooutliers.hist(color="orange",bins=10,edgecolor="black")
plt.xlabel("degree-holders Black People income after taking the training")
plt.ylabel("Frequency")
plt.xlim([0,26000])

plt.show()

# plotting the QQ plots of the income of both groups to measure the normality of the distribution
stats.probplot(white_people_group_income, plot=plt)
plt.ylabel="number of degree-holders white People income after taking the training"
plt.show()
stats.probplot(Black_people_group_income_nooutliers, plot=plt)
plt.ylabel="number of degree-holders black People income after taking the training"
plt.show()
```

Appendix B

$$t = \frac{point\ estimate - null\ value}{SE}$$

t : t-score

Point estimate $= \bar{x}_{white} - \bar{x}_{Black}$

$\bar{x}_{white}$: The mean of the income of the white people group

$\bar{x}_{Black}$: The mean of the income of the black people group.

Standard Error $\left(SE_{\bar{x}_{white} - \bar{x}_{Black}}\right) = \sqrt{\frac{\sigma_{white}^2}{n_{white}} + \frac{\sigma_{black}^2}{n_{black}}}$

σ_{white}: Standard deviation of the income of the white group sample.

σ_{black}: Standard deviation of the income of the black group sample.

n_{white}: the sample number of the white group.

n_{black}: the sample number of the black group.

Confidence interval: $point\ estimate\ \pm\ t\ \times SE$

t : t-score with the least degree of freedom between both groups

$t = \frac{point\ estimate - null\ value}{SE}$

$Degree\ of\ Freedom\ (df) =\ n - 1$

```
# calculating the a one_sided 95% confidence interval of the differences between the two means of the income of black and white people
means_difference= white_mean - black_mean # calculate the difference between the two mean
# this will be the point estimate
standard_error= np.sqrt((std_white**2)/n_white + (std_black**2)/n_black)# calculate the standard error
t_score= np.abs(white_mean-black_mean)/standard_error # calculate the t-score
df= min((n_black-1), (n_white-1))# calculate the degree of freedom
t_value= stats.t.ppf(1-(1-0.95)/2,df) #setting the level of confidence
# because we are doing one-sided test, we will not multiply it by 2
print("the confidence interval is" + "[" +str(means_difference+t_value*-int(standard_error))+"," +str(means_difference+t_value*standard_error)+"]" )
print("the t_value is " + str(t_score))
```

Appendix C

```
means_difference= white_mean - black_mean # calculate the difference between the two mean
standard_error= np.sqrt((std_white**2)/n_white + (std_black**2)/n_black)# calculate the standard error
t_score= np.abs(means_difference)/standard_error # calculate the t-score
p_value= 1- (stats.t.cdf(t_score,df)) # calculate the p-value
print("the p_value is " + str(p_value))
```

Appendix D

$$Cohen's\ d = \frac{x_{white} - x_{black}}{\sqrt{\frac{(n_{white} - 1) \times \sigma_{white}^2 - (n_{black} - 1) \times \sigma_{black}^2)}{n_{white} + n_{black} - 2}}}$$

$$Hedge's\ g = Cohen's\ d\ \times (1 - \frac{3}{4(n_{white} + n_{black}) - 9})$$

$\bar{x}_{white}$: The mean of the income of the white people group

$\bar{x}_{Black}$: The mean of the income of the black people group.

σ_{white}: Standard deviation of the income of the white group sample.

σ_{black}: Standard deviation of the income of the black group sample.

n_{white}: the sample number of the white group.

n_{black}: the sample number of the black group.

```
# Calculate the pooled Standadrd deviation, cohens d and hedges d
SD_pooled = np.sqrt((std_black ** 2 + std_white ** 2) / 2)
cohen_d = ((means_difference) / SD_pooled)
Hedge_g = cohen_d * (1 - (3 / (4 * (n_white + n_black) - 9)))
print(" hedges'g is " + str(Hedge_g))
```

Appendix E

```
# calculate the power of the test
t_score=stats.t.ppf(0.9,16)# determine the T-score which represents the 90% significance level
#  the T-score will be used to mark the rejection regions
mean_1= 0 # the mean of the null hypothesis
sd_1=1429.94 # the standard deviation of the null hypothesis
x_axis= np.arange(mean_1+sd_1*-4,mean_1+sd_1*4,0.001)# setting the x-axis to extend by 4 standard deviations from the mean
plt.plot(x_axis, norm.pdf(x_axis, mean_1, sd_1)) # plotting the curve
mean_2= -2884.50911020893 # the mean of the alternative hypothesis
sd_2=1429.9477100360712 # the standard deviation of the alternative hypothesis
# the same as the null because they are the same graph but shifted
plt.axvline(sd_1*t_score, color='r', linestyle='--',linewidth=1) # drawing a line that will present the rejection region
x_axis_2= np.arange(mean_2+sd_2*-4,mean_2+sd_2*4,0.001) # setting the x-axis to extend by 4 standard deviations from the mean
px=np.arange(-8000,sd_1*t_score,0.001) # defining the shaded region
iq=stats.norm(mean_2,sd_2)
plt.fill_between(px,iq.pdf(px),color = "orange") # shading the area which represents the power of the test
plt.plot(x_axis_2, norm.pdf(x_axis_2, mean_2, sd_2))
plt.xlabel("Mean of the difference between the income of Black people and White people")
# plt.show()

power= ((sd_1*t_score) - mean_2)/sd_1
print("The power of the test is" + str(stats.t.cdf(power,44+17-2)))  # calculating the power of the test
```