

Anupam Mazumdar

Welsh Trinity Saint David, London

04/01/2021

Data Analysis & Visualisation (ACCB4002_LON)

Introduction

This analysis is based on the data which has been presented in a recent BBC (British Broadcasting Corporation) news by specifying that the impact of covid-19 is different among the white people of UK (United Kingdom) and the BAME (black, Asian and minority ethnic) people of UK. The age of 50 white patients and BAME patients from the UK has been taken for the analysis whose age is between the intervals of 20 to 50. The basic aim of this analysis is to determine the ratio of infection rate and the truthfulness of these findings that either the data of these 50 patients are accurate and reliable. The main factors which are going to be addressed in this analysis are the development of an appropriate set of testable hypotheses, then test that hypothesis to the $p=0.10$, $p=0.01$, and $p=0.02$ significance level. The analysis also determines any presence of outliers in that data and their impact on the outcomes.

Developing Hypotheses

The three testable hypotheses for this experiment are defined below:

- Null Hypothesis: H_0 : If the $p = 0.10$, then the data is insignificant.
- Alternative Hypothesis: H_1 : If the $p < 0.10$, then the data is significant.
- Null Hypothesis: H_0 : If the $p = 0.01$, then the data is insignificant.
- Alternative Hypothesis: H_1 : If the $p < 0.01$, then the data is significant.
- Null Hypothesis: H_0 : If the $p = 0.02$, then the data is insignificant.
- Alternative Hypothesis: H_1 : If the $p < 0.02$, then the data is significant.

Testing the Hypotheses

The appropriate technique which has been used to perform this analysis is the t-test by utilizing the data analysis tool pack. For testing the hypothesis to a significance level of $p=0.10$, $p=0.01$, and $p=0.02$, the t-test has been used for the two samples of White Patients and BAME assuming the unequal variances. The result of the significance level of $p = 0.10$ is defined below:

	White Patient	BAME
Mean	36.18	36.74
Variance	88.72204082	86.80857
Observations	50	50
Hypothesized Mean Difference	0	
df	98	
t Stat	-0.298879822	
P(T<=t) one-tail	0.382832305	
t Critical one-tail	1.290249904	
P(T<=t) two-tail	0.76566461	
t Critical two-tail	1.660551217	

For testing the hypothesis, data of both Covid-19 infected white patients and BAME patients has been analyzed and the results of this analysis can be interpreted as the p-value is 0.7656 which is greater than the significance level which is 0.10, so there is no evidence of rejecting the null hypothesis H_0 and accepting the alternative hypothesis H_1 . It has been concluded that when the $p = 0.10$, then the data is insignificant. The same hypothesis can also be tested by assuming the t-value of this output. The results can be interpreted as because the value of t statistics is -0.298 which is smaller than the t-value of the two-tail test which is 1.6605, so there is no evidence to reject the null hypothesis H_0 . It has been concluded that when the $p = 0.10$, then the data is insignificant. It has evidently shown that from both perspectives it has been concluded that do not reject the null hypothesis.

The result of the significance level of $p = 0.01$ is mentioned below:

	White Patient	BAME
Mean	36.18	36.74
Variance	88.72204082	86.80857143
Observations	50	50
Hypothesized Mean Difference	0	
df	98	
t Stat	-0.298879822	
P(T<=t) one-tail	0.382832305	
t Critical one-tail	2.36500241	
P(T<=t) two-tail	0.76566461	
t Critical two-tail	2.626931096	

For testing the hypothesis, data of both Covid-19 infected white patients and BAME patients has been analyzed and the results of this analysis can be interpreted as the p-value is 0.7656 which is greater than the significance level which is 0.01, so there is no evidence of rejecting the null hypothesis H_0 and accepting the alternative hypothesis H_1 . It has been concluded that when the $p = 0.01$, then the data is insignificant. The same hypothesis can also be tested by assuming the t-value of this output. The results can be interpreted as, the value of t statistics is -0.298 which is smaller than the t-value of the two-tail test which is 2.626, so there is no evidence to reject the null hypothesis H_0 . It has been concluded that when the $p = 0.01$, then the data is insignificant. From both perspectives, it has been concluded that do not reject the null hypothesis.

The result of the significance level of $p = 0.02$ is shown below:

	White Patient	BAME
Mean	36.18	36.74
Variance	88.72204082	86.80857
Observations	50	50
Hypothesized Mean Difference	0	
df	98	
t Stat	-0.298879822	
P(T<=t) one-tail	0.382832305	
t Critical one-tail	2.081444921	
P(T<=t) two-tail	0.76566461	
t Critical two-tail	2.36500241	

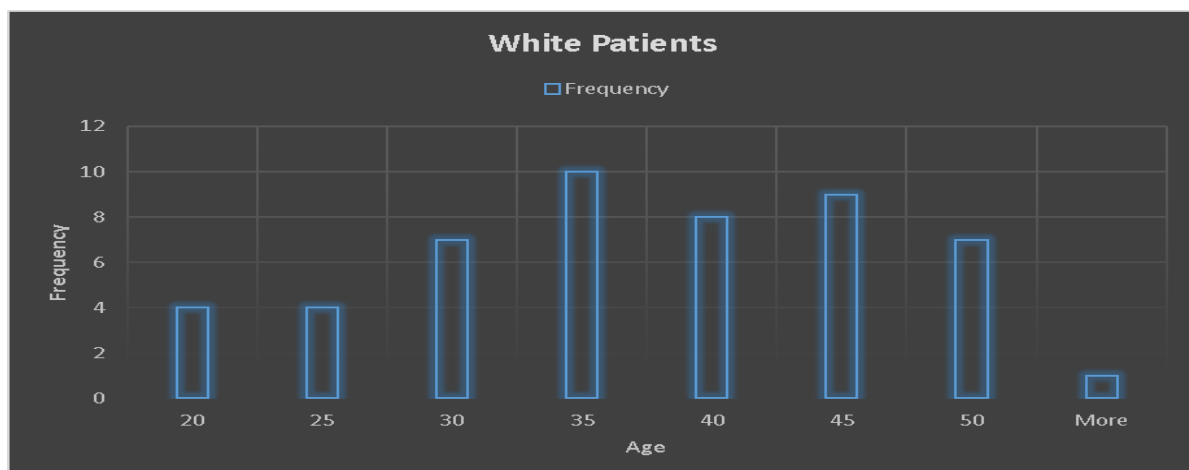
After analyzing the data of both Covid-19 infected white patients and BAME patients the above-mentioned statistics have been obtained and the results of this analysis can be interpreted as the p-value is 0.7656 which is greater than the significance level of 0.02, so there is no manifest evidence to reject null hypothesis H_0 and to accept the alternative hypothesis H_1 . It has been concluded that when the $p = 0.02$, then the data is insignificant. The hypothesis can be tested again by assuming the t-value of this output. The results of the t-value can be

interpreted as, the value of t statistics is -0.298 which is smaller than the t-value of the two-tail test which is 2.365, which identifies that there is no evidence to reject the null hypothesis H_0 . It has been concluded that when the $p = 0.02$, then the data is insignificant. From both perspectives, it has been concluded that do not reject H_0 . After testing the hypotheses to a significance level of $p=0.10$, $p=0.01$, and $p=0.02$, it has been observed that the conclusion is the same for all the significance level. Either it is tested by p-value or by t-value both have given the same conclusion that does not reject H_0 .

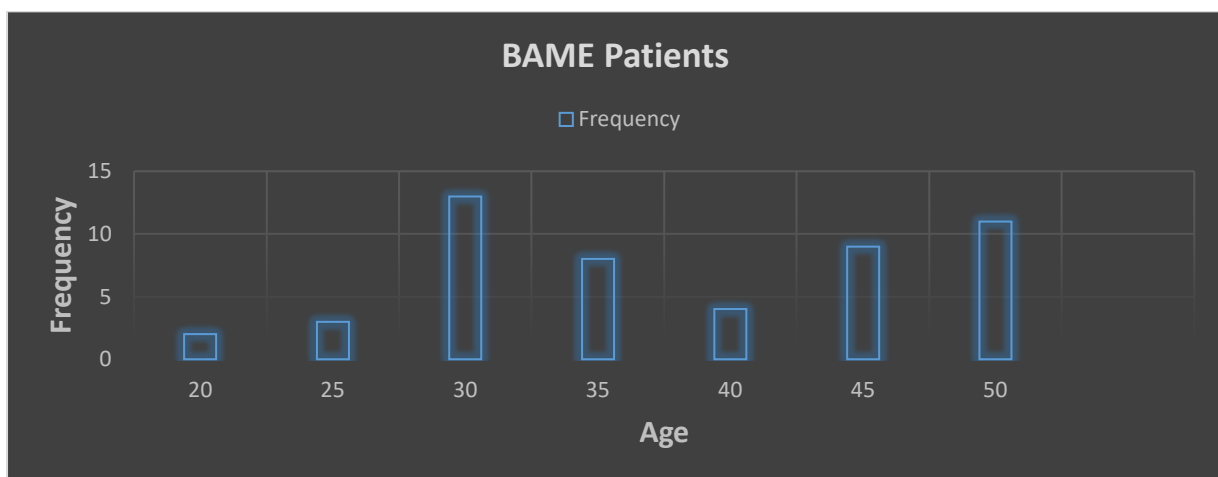
Presence of outliers

There is an outlier in the data that the age of white people of the UK who are infected with covid-19 is supposed to be between the intervals of 20-50, but the age of the 9th patient from this data is greater than this interval. One of the white patients is 55 years old who get infected by covid-19. Although all of the patients of BAME has an age between the intervals of 20-50. However, this change of age in the white patient has affected the overall outcome of this analysis because it has assumed that the ages of both of the patients either it is white or BAME are equal, and now when an outlier is found in the age of patients so it caused problems in this statistical analysis and made the outcomes of this data less efficient and less competent.

Features of data



The other features of data which has been appeared during this analysis are that out of the white people who got infected with covid-19 the ratio of infection rate is high in the people whose age is around 35 years. It has been analyzed through the graphical representation of the dataset, seven of the white patients has around 50 years of age, nine of the white patients has around 45 years of age, 8 of the patients has around 40 years of age, 7 of the patients has around 30 years of age, 4 of the patients has around 25 years of age and 4 of the patients has around 20 years of age. And there is a white patient whose age is more than the selected interval, the patient has 55 years of age. It has been observed that after the age of 25 the ratio of infection rate has been increased in the white patients and its effects have increased as the age of the patient increased.



It has been observed after analyzing the data set of BAME patients that the ratio of infection rate of Covid-19 is high in the people who have age around 30 years and the people whose age is between the intervals of 20 to 29 has very fewer chances to get affected by this infection. Eleven of the BAME patients has around 50 years of age, nine of the white patients has around 45 years of age, four of the patients has around 40 years of age, eight of the patients has around 35 years of age, thirteen of the patients has around 30 years of age, three of the patients has around 25 years of age and just two of the patients has around 20 years of age. So, the feature

of data that has been appeared during the investigation of the dataset is that the ratio of infection rate is high in the patients whose age is beyond 30 years.

Other factors that might affect results

The other factors which might affect the results but have not been considered in the above-stated analysis are the factors of white and BAME's people outdoor activities and literacy. Such as, the people who get infected by this infection are mostly because of this they are not following the proper SOPs (Standard operating procedure) they are not doing proper sanitation and not conforming the proper rules before leaving their houses (Bhardwaj, 2020). The infection rate also diverges based on the factor of outdoor activities like the people who are engaged more in the outdoor activities have a high probability to get infected as compared to the people who live inside the house and do not have many activities due to which they leave the house. These factors can be controlled by proper awareness in people about this infection that how important is to have the proper precautions for getting prevented from covid-19.

Conclusion

In the concluding statements, it can be said that the two-tail test technique had been used for performing this analysis. The analysis determined that the hypotheses of the significance level of $p=0.10$, $p=0.01$, and $p=0.02$, have been failing in rejecting the null hypothesis H_0 which concludes that the data is insignificant.

References

“Bhardwaj, A. (2020). COVID-19: Challenges & Its Adaptation. Available at SSRN 3659239”.