



Data Analysis and Visualization

(ACCB4002)

Assessment 2

Student:
Ramona Georgiana Ghinea
2009126

Lecturer:
Anupam Mazumdar

Date: 06.01.2021

Table of Contents

Introduction.....	1
Dataset Description	3
Appropriate Hypotheses for Analysis	3
<i>F-test two sample for variance:.....</i>	<i>3</i>
<i>Selection of Hypotheses:</i>	<i>4</i>
<i>T-test for unequal variances:</i>	<i>4</i>
Hypothesis.....	4
Null Hypothesis	4
Alternative Hypothesis	4
Hypotheses for Different Significance Level	5
A) $P = 0.10$	5
B) $P = 0.01$	6
C) $P = 0.02$	6
Application of techniques used.....	7
Insights and Visualizations from the Data	8
Scatter plot	8
Box plot Bar graph:.....	8
Bar graph:	9
Results	9
Conclusion	9
References.....	10

Introduction

This report consists of testing of findings of the BBC news of UK, they identified that the infection rates of Covid-19 of White and BAME patients of UK is varying and the hypothesis is made on the basis of that. This report includes the t-test for unequal variances which is to be performed on the Data. Graphs, applications and results are been discussed in this report.

Dataset Description

The Dataset consist of UK news data, related to the infection rate of Covid-19. The ages of BAME and White patients is been given. 50 patient's data from both the categories are there and the age of these patients is given in the Dataset. The age lies in the range of 20 to 50 years. The patient number is there which shows the primary key for the data. The infection rate of covid-19 and in order to find out the truthfulness in the findings the hypotheses analysis will be done.

Appropriate Hypotheses for Analysis

F-test two sample for variance:

This test was done in order to calculate the variances of two samples and so the hypotheses test can be processed.

F-Test Two-Sample for Variances		
	<i>Variable 1</i>	<i>Variable 2</i>
Mean	36.18	36.74
Variance	88.72204082	86.80857143
Observations	50	50
df	49	49
F	1.022042402	
P(F<=f) one-tail	0.469740831	
F Critical one-tail	1.607289463	

Figure 1 f-test to measure variance Figure 1 f-test to measure variance

From this test it can be seen that the variances for bot of the variables is different and so the test for the unequal variances must be followed (Steinberger 2016).

Selection of Hypotheses:

T-test for unequal variances:

This is the hypotheses which can be done on two independent population which are having unequal variances. This test of two types, one-tail and two-tail t-test. Here the variances of both the variables are different so the t-test for unequal variances can be performed (Wang et al. 2016).

Hypothesis

Hypothesis is an approximate value which can be tested for investigation and testing can be done to find the truthfulness of the data. There are two types of hypothesis which are null hypothesis and alternate hypothesis. They are useful in research purpose they provide clarification to the research.

Null Hypothesis

Null Hypothesis is denoted as H_0 , if it is rejected then it states the opposite of what the researcher has expected or predicted. It tells the user that there is no actual relationship between the variables. Null hypothesis is established when, $\mu_1 - \mu_2 = 0$

Alternative Hypothesis

It is denoted as H_1 , it states the same result the researcher is expecting, it is categorized into two categories. Directional and non-directional hypothesis. When the difference of means is not equal to zero. Then the alternative hypothesis is established (Statisticssolutions n.d.).

Hypotheses for Different Significance Level

A) $P = 0.10$

For the significance level 0.10, the test is performed and found the following result. In which the probability value is found out in the results, the t-critical values. The Analysis of the result can be done in following ways:

Step 1:

Null hypothesis cannot be rejected if the value of t stat is smaller than the value of t critical one tail. Here from the observation, it can be seen that the t stat value is -0.298 is smaller than 1.660 then the null hypothesis cannot be rejected and the infection rates between the BAME and White patients is varying as it was said by the BBC news of UK.

Step 2:

In this step the p-value is taken in consideration, if the value of p is greater than alpha which is 0.10, and the p-value for two tail is 1.666 then the Null hypothesis cannot be rejected (Excel-easy n.d.).

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
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	1.290249904	
P(T<=t) two-tail	0.76566461	
t Critical two-tail	1.660551217	

Figure 2 t-test for $p=0.10$

B) $P = 0.01$

For the significance value 0.01, the result is analyzed:

Step 1: If the t-stat value is less than the t-critical values from the observation it can be seen that the null hypothesis cannot be rejected.

t-stat value = -0.298

t critical two tail value = 2.626

It is clearly seen that the t-stat value is less than the t critical value and thus the Null hypothesis cannot be rejected.

Step 2: From the result, it can be seen that the p-value for two tail is greater than the significance value then the null hypothesis cannot be rejected, thus from the both cases it is seen that the Null hypothesis which is made cannot be rejected (Excel-easy n.d.).

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
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	

Figure 3 t-test for $p=0.01$

C) $P = 0.02$

At the significance value of 0.02, the values can be analyzed as:

Step 1:

The t-stat value from the observation is: -0.298

The t critical two tail value: 1.660

It is seen that the t-stat value is less than the t critical value then the Null hypothesis cannot be rejected.

Step 2:

Here, P-two tail value: 0.765

Significance value = 0.02

It is clearly seen that the p-two tail value is greater than the significance value. And null hypothesis is hard to reject (Excel-easy n.d.).

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	Variable 2
Mean	36.18	36.74
Variance	88.72204082	86.8086
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	

Figure 4 t-test $p=0.02$

Application of techniques used

1. It is used to test variety of population on which some assumptions are made.
2. By using the hypothesis, the credibility of the sample data can be checked.
3. It provides the researcher whether the assumption he/she has made is correct or not.
4. Random samples and big sample regarding any population can be tested using the hypothesis.
5. T-test can be used for equal and unequal variances, and various statistical problems can be solved using this (Majaski 2020).

Insights and Visualizations from the Data

Scatter plot

The scatter plot shows the two populations at various distributed points.

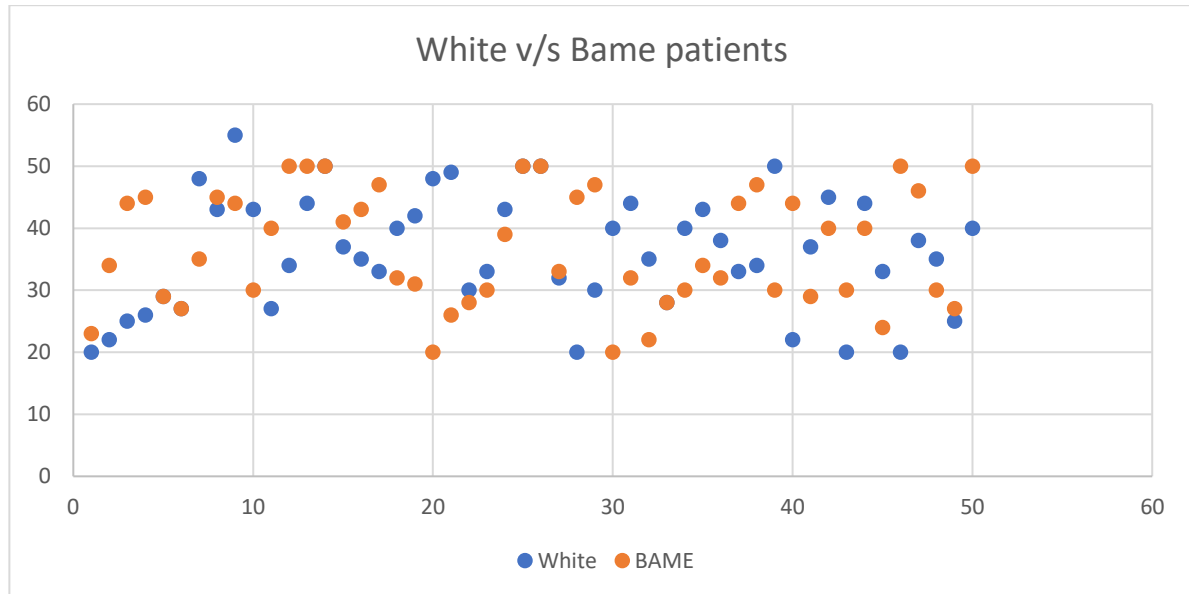


Figure 5 Scatter plot

Box plot Bar graph:

Box plot shows the Data in box form of both the WHITE and BAME patients.

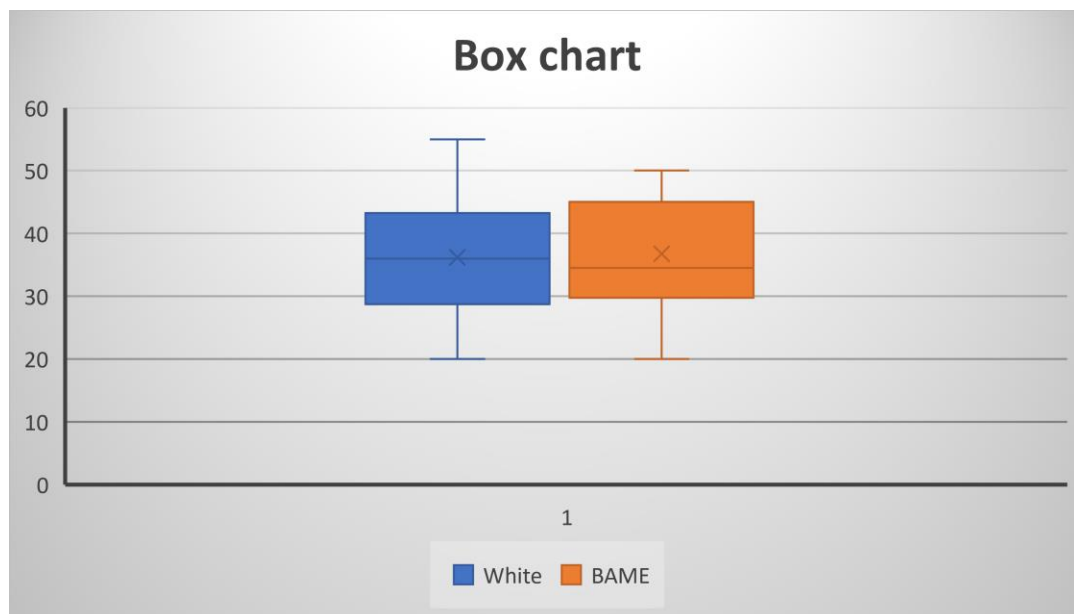


Figure 6 Box plot

Bar graph:

The bar graph shows the Data of Various patients at different place.

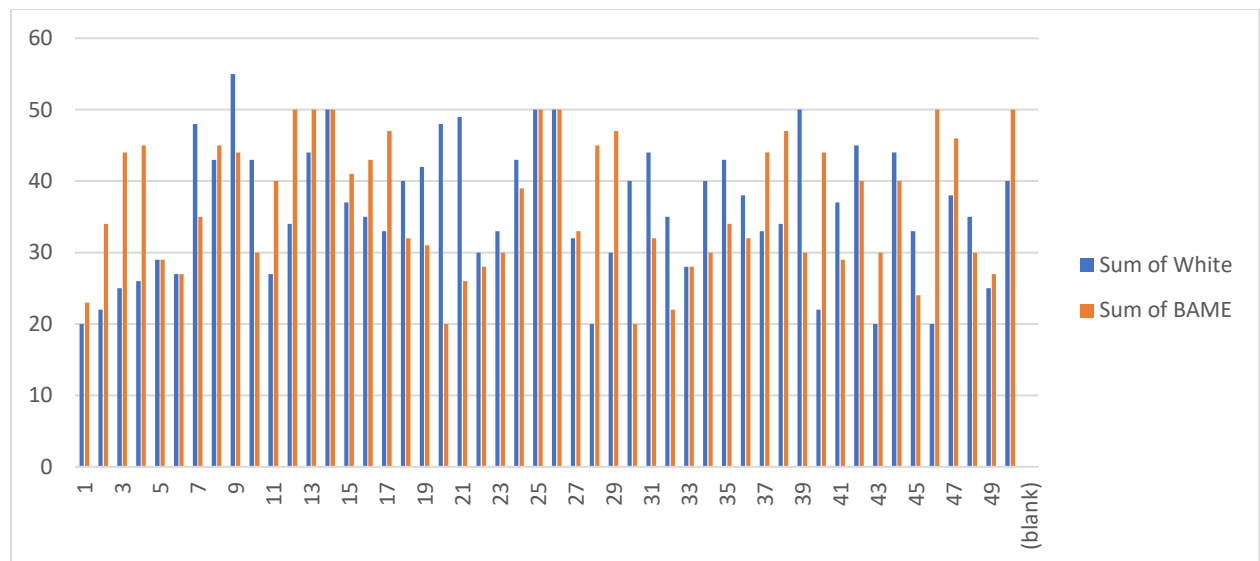


Figure 7 Bar graph

Results

From the observation it can be seen that the Null hypothesis in all the significance levels at $p=0.10$, 0.01 , 0.02 and it is found that the Null hypothesis cannot be rejected in all the cases so the assumption which was made as the infection rates in the BAME and White patients is varying is true.

Conclusion

In this report, the hypothesis testing is done using the Data analysis function of excel, the Data of patients of UK of two community BAME and White of infection of COVID-19, the t-test for various significance level was done. And appropriate graphs and the application of appropriate technology was discussed.

References

Excel-easy. N.d. t-Test. Excel-easy, [online], Available at: <https://www.excel-easy.com/examples/t-test.html>. [Accessed on 22 December 2020]

Majaski, C. 2020. Hypothesis testing. Investopedia, [online], Available at: <https://www.investopedia.com/terms/h/hypothesistesting.asp#:~:text=Key%20Takeaways-,Hypothesis%20testing%20is%20used%20to%20assess%20the%20plausibility%20of%20a,o%20the%20population%20being%20analyzed.> [Accessed on 22 December 2020]

Statisticssolutions. N.d. Null hypothesis and Alternative Hypothesis. Statisticssolutions, [online], Available at: https://www.statisticssolutions.com/null-hypothesis-and-alternative-hypothesis/?__cf_chl_jschl_tk__=123fb97f9ee1c8a668974b4798c007e6d457940e-1608614950-0-AcclyjtqVhcmupUWA2B7HceUaL_nlfcl_xlEgUBhksi1b7Rga-6fK3CjqFiZ9JlIhzAr8gkQ-LKlo-lblR9XRTEKMupwzantFie_1ED9s__r0wLEL3TZgEK3ju5eOdmDeFdXeFEj6DX8gxKUndWhFew6vevD54Dzay709mkS_QFLyRBrqLAyELPVp8Y6aFzBeU89sVEAqfii40lokFCHxBg3HErTuWPuG1bRGU7pbDuUzYucY_234q1njg4y9ilDuA-NTzN0GYqaM4_nCAUIqhPOavlQINq6r1F3gdzReB2D0RL3sBy9J_gWC8cy-8Y6ckyXp-UU1lQnVNMm7h1jQ0BEqJ6s7EKwbuqZG9ZlHbPZ. [Accessed 22 December 2020]

Steinberger, L., 2016. The relative effects of dimensionality and multiplicity of hypotheses on the F -test in linear regression. *Electronic Journal of Statistics*, 10(2), pp.2584-2640.

Wang, M. and Liu, G., 2016. A simple two-sample Bayesian t -test for hypothesis testing. *The American Statistician*, 70(2), pp.195-201.

