

Data Analysis and Visualization (ACCA4002)
Assignment 2

Rumyana Yaneva (2011961)
University of Wales Trinity Saint David
Lecturer: Anupam Mazumdar
Date: 29.12.2020

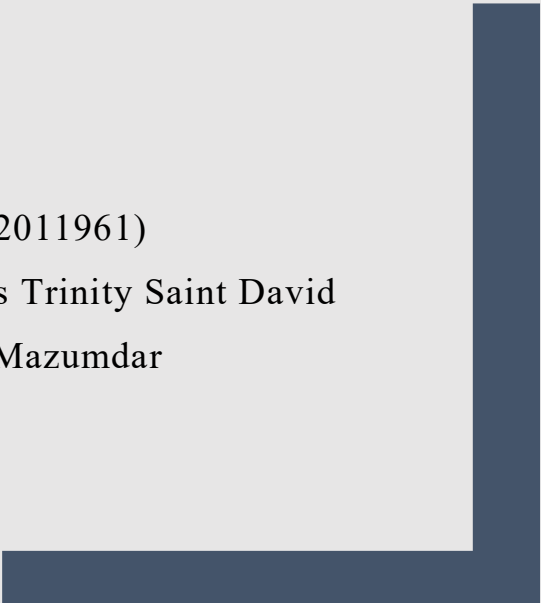


Table of Contents

Introduction	2
Main Body	2
Deployment.....	6
1.Hypothesis:	13
2.Hypothesis:	13
Conclusion	13
References.....	14

Introduction

In the given assessment task, we are asked to do analysis of the given data set and analyse the information in the recent BBC (UK) news, in which has been identified that the infection rate of Covid – 19 is varied between white people and BAME in the UK. The analysis includes a comprehensive data collection to determine the patients primarily affected by the COVID-19 outbreak. Based on the data we are required to analyse the ratio of infection rate and truthfulness of the findings. To do this, a variety of hypotheses are optimised to be tested on the basis of relevant mathematical techniques.

Main Body

Analysis

Description of the data:

We have 50 different white and BAME patients all aged between 20 and 50 years old. The evidence in the report shows the omnipresence of infectious illnesses of various groups of white and BAME people. The purpose of the data collection is to determine the infection frequency ratios of all demographic groups, especially White & BAME. The succinct evaluation of the descriptions is dealt with here in the following direction:

Descriptive statistics

	White	BAME
Average	36.18	36.74
Patients Number	50	50
Median	36	34.5
Mode	20	50
St. Deviation	9.4192	9.3171
Variance	88.722	86.809

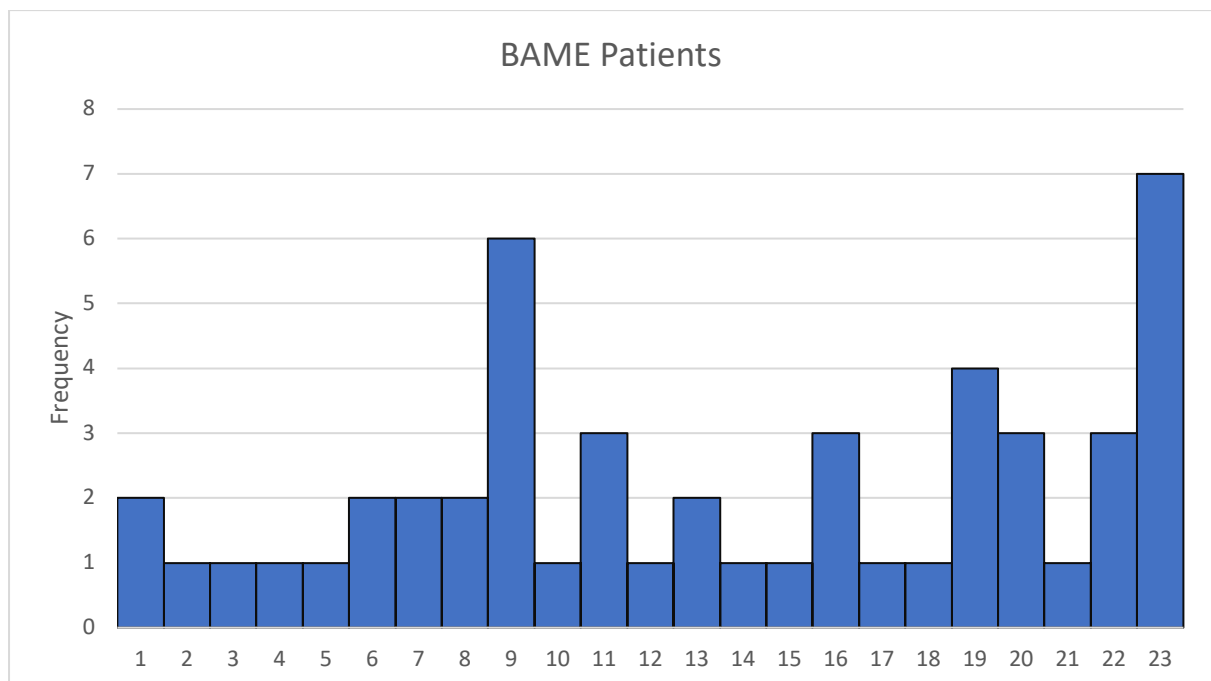
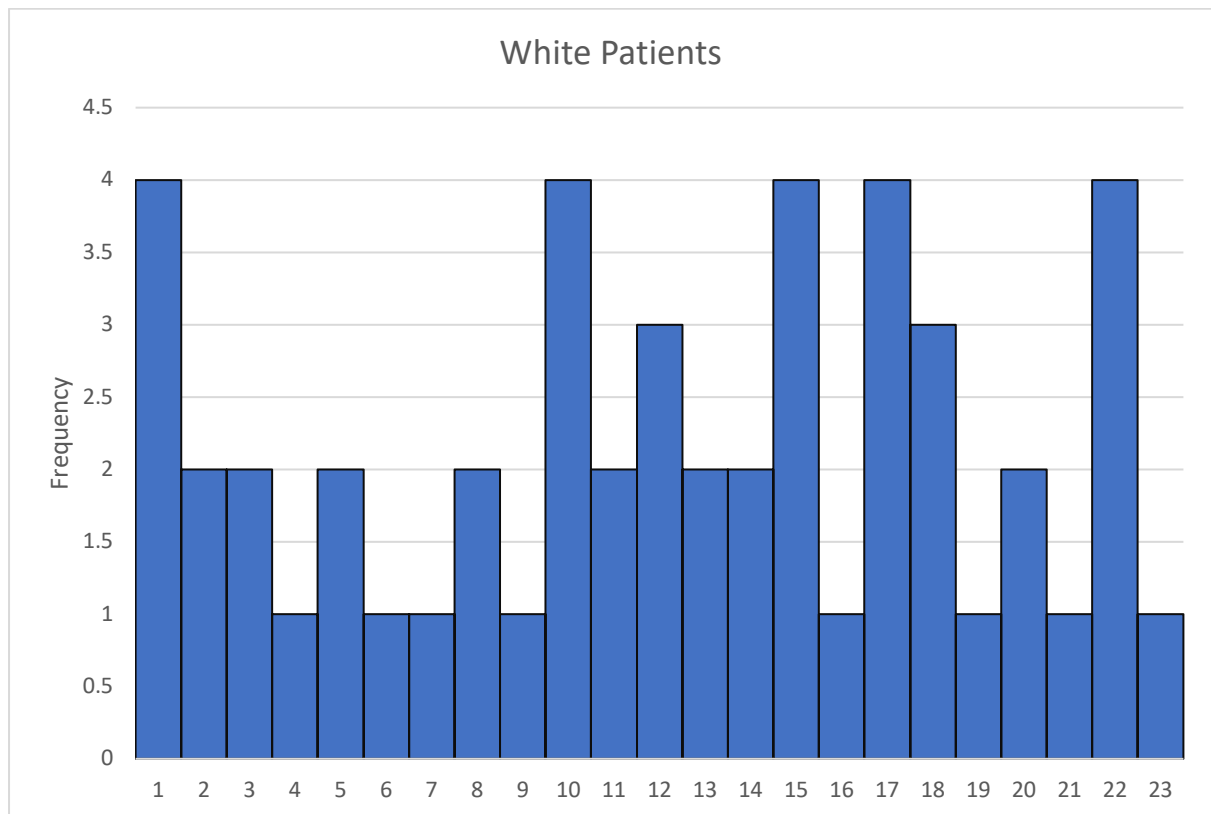
a. There exist several modes. The lowest value that is seen is

White Patients

Patients Class	Frequency	Relative Frequency	Percent frequency	Cumulative Frequency
20	4	0.08	8	8
22	2	0.04	4	12
25	2	0.04	4	16
26	1	0.02	2	18
27	2	0.04	4	22
28	1	0.02	2	24
29	1	0.02	2	26
30	2	0.04	4	30
32	1	0.02	2	32
33	4	0.08	8	40
34	2	0.04	4	44
35	3	0.06	6	50
37	2	0.04	4	54
38	2	0.04	4	58
40	4	0.08	8	66
42	1	0.02	2	68
43	4	0.08	8	76
44	3	0.06	6	82
45	1	0.02	2	84
48	2	0.04	4	88
49	1	0.02	2	90
50	4	0.08	8	98
55	1	0.02	2	100
Total	50		100	

BAME Patients

Patients Class	Frequency	Relative Frequency	Percent Frequency	Cumulative Frequency
20	2	0.04	4	4
22	1	0.02	2	6
23	1	0.02	2	8
24	1	0.02	2	10
26	1	0.02	2	12
27	2	0.04	4	16
28	2	0.04	4	20
29	2	0.04	4	24
30	6	0.12	12	36
31	1	0.02	2	38
32	3	0.06	6	44
33	1	0.02	2	46
34	2	0.04	4	50
35	1	0.02	2	52
39	1	0.02	2	54
40	3	0.06	6	60
41	1	0.02	2	62
43	1	0.02	2	64
44	4	0.08	8	72
45	3	0.06	6	78
46	1	0.02	2	80
47	3	0.06	6	86
50	7	0.14	14	100
Total	50		100	



Development of appropriate hypotheses:

The premise of an analysis is that the possible consequences of a human population science study, like the supposed discrepancies in findings across populations on a specific individual or allegiances, are real, specific and provable statements of mathematical theories.

H0: There is a major difference in the incidence of viral infections in whites and BAMEs.

H1: There is no major difference in the incidence of viral infections in whites and BAMEs.

H0: There is strong relationship between the occurrence of BAMEs and Whites

H1: There is no relationship between the occurrence of BAMEs and Whites

Deployment

Selected analyses techniques

There are a whole range of techniques and practises of conducting an evaluation, and yet, in a sequence of measurements, the ANOVA test is among the valuable tools for evaluating a substantial variability. In addition to this, the analysis of correlation is an essential technique for evaluating the correlation between white and BAME research analysis.

- One sample T-test: with one t-testing, if the data obtained is completely opposite from any significance, it is used to validate. All gives idea on how a certain degree of significance M is due to error bars reflecting μ below. Also, t-test is a contemporary method of comprehensive analysis that is used to assess if the different criteria that might have been added to those structures have a significant difference. If selected databases behave as directed, it is used and may have unexplained variations, such as with reported information source as a result of tossing some dice 75 times. A t-test may be used as a thesis test tool, which involves group-specific statement testing (Yang, Chou and Chen, 2020).
- Correlation: In statistical measures, association or density is any causal relationship, either mutual or semi, among connection between two variables or semi data. In the highest potential meaning, connection is any clear relationship, but it typically refers to the degree by which a set of requirements is concurrently connected. Coefficients are useful because it is possible to assume a correlation that can be skewed in nature. For example, demonstrate the relation with electricity production and heat that the electricity company would generate less electricity on a normal week. Apparently, in this case, there appears to be a possible correlation, as poor weather encourages us to

provide more energy for heating systems. In fact, the essence of a partnership is not likely to be necessary to include a semi connection.

Techniques to present the results:

Through the section, there seems to be no systematic method or tool to illustrate the analysis of evidence. It is also necessary to assess the provision of report and charts for the visualisation of comprehensive observations from various kinds of studies. According to the above portion, including appropriate tables and graphs in analytics and insights, two modes of study, Regression analysis and similarities, will be presented just so researchers also might understand the real nature of the bacterial occurrence between whites and BAME in Great Britain. For a detailed analysis of the rate of inflation of whites and BAME, below are a selection of estimates and data.

Testing of the hypotheses to significance level of $p=0.10$, $p=0.01$, and $p=0.02$

Significance level of $p=0.10$

One-Sample Statistics

		Statistic	Bootstrap ^a			
			Bias	Std. Error	90% Confidence Interval	
					Lower	Upper
White patient	N	50				
	Mean	36.180	.017	1.309	33.941	38.239
	Std. Deviation	9.4192	-.1335	.7203	8.1647	10.4345
	Std. Error Mean	1.3321				
BAME	N	50				
	Mean	36.740	-.009	1.333	34.441	38.880
	Std. Deviation	9.3171	-.1038	.5523	8.3032	10.1280
	Std. Error Mean	1.3176				

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
White patient	27.161	49	.000	36.1800	33.947	38.413
BAME	27.883	49	.000	36.7400	34.531	38.949

Bootstrap for One-Sample Test

	Mean Difference	Bootstrap ^a				
		Bias	Std. Error	Sig. (2-tailed)	90% Confidence Interval	
					Lower	Upper
White patient	36.1800	.0175	1.3088	.001	33.9410	38.2390
BAME	36.7400	-.0092	1.3328	.001	34.4410	38.8800

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Significance level of $p=0.01$ **One-Sample Statistics**

		Statistic	Bootstrap ^a			
			Bias	Std. Error	90% Confidence Interval	
					Lower	Upper
White patient	N	50				
	Mean	36.180	.029	1.322	34.041	38.499
	Std. Deviation	9.4192	-.1412	.7189	8.1234	10.4846
	Std. Error Mean	1.3321				
BAME	N	50				
	Mean	36.740	-.016	1.314	34.561	38.820
	Std. Deviation	9.3171	-.1231	.5529	8.2384	10.0995
	Std. Error Mean	1.3176				

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	99.9% Confidence Interval of the Difference	
					Lower	Upper
White patient	27.161	49	.000	36.1800	31.517	40.843
BAME	27.883	49	.000	36.7400	32.128	41.352

Bootstrap for One-Sample Test

	Mean Difference	Bootstrap ^a				
		Bias	Std. Error	Sig. (2-tailed)	90% Confidence Interval	
					Lower	Upper
White patient	36.1800	.0294	1.3218	.001	34.0410	38.4990
BAME	36.7400	-.0160	1.3143	.001	34.5610	38.8200

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Significance level of $p=0.02$ **One-Sample Statistics**

		Statistic	Bootstrap ^a			
			Bias	Std. Error	90% Confidence Interval	
					Lower	Upper
White patient	N	50				
	Mean	36.180	.004	1.258	34.140	38.299
	Std. Deviation	9.4192	-.1164	.6726	8.1665	10.4398
	Std. Error Mean	1.3321				
BAME	N	50				
	Mean	36.740	-.008	1.305	34.501	38.840
	Std. Deviation	9.3171	-.1215	.5635	8.1801	10.1315
	Std. Error Mean	1.3176				

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	98% Confidence Interval of the Difference	
					Lower	Upper
White patient	27.161	49	.000	36.1800	32.976	39.384

BAME	27.883	49	.000	36.7400	33.571	39.909
------	--------	----	------	---------	--------	--------

Bootstrap for One-Sample Test

	Mean Difference	Bootstrap ^a				
		Bias	Std. Error	Sig. (2- tailed)	90% Confidence Interval	
					Lower	Upper
White patient	36.1800	.0043	1.2584	.001	34.1400	38.2990
BAME	36.7400	-.0078	1.3048	.001	34.5010	38.8400

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Hypothesis 2 Correlations

		White patient	BAME
White patient	Pearson Correlation	1	.072
	Sig. (2-tailed)		.621
	N	50	50
BAME	Pearson Correlation	.072	1
	Sig. (2-tailed)	.621	
	N	50	50

Interpretation of the results:

Taking view of the above detailed analysis of the given statistical model, it can be inferred that, while the error margin is about 9.42, the mean value is 36.18. It may also be suggested

that there is very little connection with respect to the occurrence of white infection and BAME. Therefore, if the reported mean and standard deviation variables vary from the others, it may be inferred that graphs are not really available to others.

1.Hypothesis:

At the value of 0.10- On the basis of the small tail test performed above, it may be stated that the value of the variability of real meaning or P is 0.00 and less than 0.05. Therefore, it will be considered that the null hypothesis is rational, but there is a major gap between the incidence of chronic diseases in whites and BAME at the cut line of 0.10.

At the value of 0.01- Equally, often at this level of 0.01, it will be mentioned that the p - values is equal to the 0.001 level, and less than the 0.05 significance. As a result, it may be verified that the null is correct, but there is a substantial difference between white and BAME infection rates at 0.01.

At the value of 0.02 - Consequently, there is a significant difference of 0.00 at this rate of 0.02 and it is less than 0.05, which means that there must be a null hypothesis. There is also a significant great difference between occurrence of infection in white people diagnosed and the average of BAME of nearly 0.02 (Schmitz, Sekulla and Pape, 2020) (Schmitz, Sekulla and Pape, 2020).

2.Hypothesis:

In the second hypothesis, correlation has been carried out. Based on the measures identified, it should be noticed that the magnitude of Pearson correlation coefficient is about 0.72, which is greater than 0.6. It could also be reported that the occurrence of infection in whites and BAME tend to have a convincing and specific relationship. In addition, the value of the correlation difference is 0.621, which means there is no null hypothesis, because the value is slightly greater than 0.05.

Conclusion

It is concluded from the above-mentioned research that there may be a potential risk between the incidence of white infection and BAME. There is indeed a common connection between both the occurrence of white and BAME. It also shows that the connection in infection rate is similar to all the other portion between both white & BAME.

References

- Kranzler, J.H. (2018). Statistics for the terrified, Rowman & Littlefield, New York (USA)
- Kruse, K., Hug, C.B. and Vaquerizas, J.M., 2020. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biology*, 21(1), pp.1-19.
- Yang, S.Y., Chou, W.H. and Chen, H.X., 2020. Development of the automated tool to perform analysis and visualisation of user reviews. *International Journal of Arts and Technology*, 12(3), pp.218-237.
- Schmitz, C., Sekulla, A. and Pape, S., 2020, June. Asset-Centric Analysis and Visualisation of Attack Trees. In *International Workshop on Graphical Models for Security* (pp. 45-64). Springer, Cham.