



DATA ANALYSIS & VISUALISATION
ACCB4002_LON
GEORGI KARACHOROV
ID 2011240
ASSESSMENT 2

This is Assessment 2 from Data Analysis and Visualisation October 2020 Term 1.

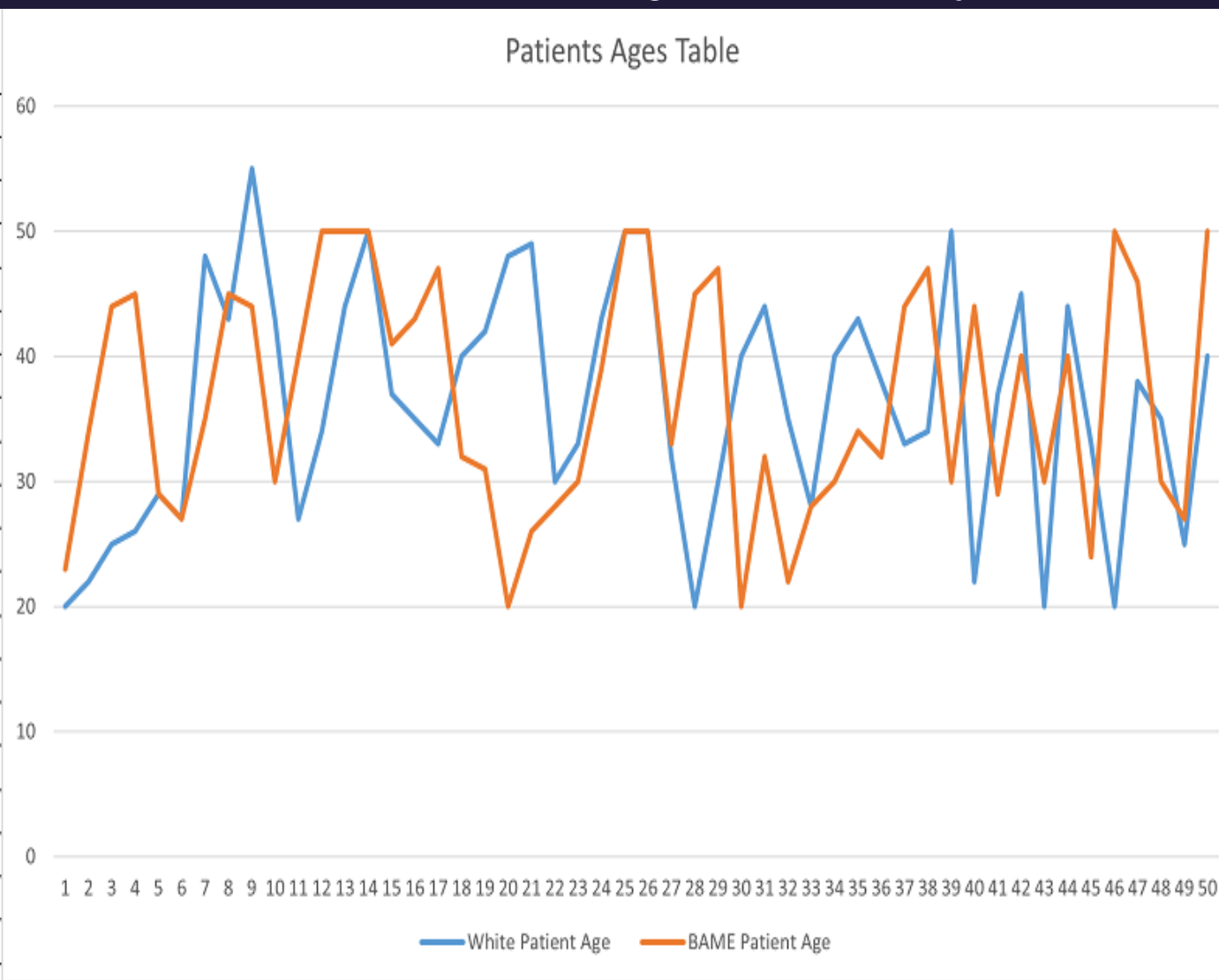
A report about the infection rate of Covid-19 between white and BAME people in the UK, based on their ages.

In this report I will challenge my analysing skill, to provide information about the data that is showed in the next slide and give as best as possible understanding and truthfulness.

To avoid high similarity, I will not use analytic words as: hypothesis, value, average and so on.

The data is sample of 50 random people, from both white and BAME races, between age of 20 and 50 years old

Patient No	White Patient	BAME	Patient No	White Patient	BAME
1	20	23	26	50	50
2	22	34	27	32	33
3	25	44	28	20	45
4	26	45	29	30	47
5	29	29	30	40	20
6	27	27	31	44	32
7	48	35	32	35	22
8	43	45	33	28	28
9	55	44	34	40	30
10	43	30	35	43	34
11	27	40	36	38	32
12	34	50	37	33	44
13	44	50	38	34	47
14	50	50	39	50	30
15	37	41	40	22	44
16	35	43	41	37	29
17	33	47	42	45	40
18	40	32	43	20	30
19	42	31	44	44	40
20	48	20	45	33	24
21	49	26	46	20	50
22	30	28	47	38	46
23	33	30	48	35	30
24	43	39	49	25	27
25	50	50	50	40	50



BBC's records about the infection of Covid-19 between white and BAME people in the UK

When we look to this data set, it looks like there is not big difference between the two groups, but we can not give any conclusion till analytic evidences and calculation have been provided. To do so I will perform T-test to find out if whether there is any different between those two group of people. Due to the small amount of this data, I will use an equal variances test. I will use Microsoft Excel for the purpose. The information that I want to find, is which group has more young infected people and if there is any big difference between those two groups.



	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		<i>Variable 1</i>	<i>Variable 2</i>
4	Mean	36.18	36.74
5	Variance	88.722	86.8086
6	Observations	50	50
7	Pooled Variance	87.7653	
8	Hypothesized Mean Difference	0	
9	df	98	
10	t Stat	-0.29888	
11	P(T<=t) one-tail	0.38283	
12	t Critical one-tail	1.66055	
13	P(T<=t) two-tail	0.76566	
14	t Critical two-tail	1.98447	

This is the result from my t-T. Now we have clear view about the information of this data set. I have set up my alpha P. on value of 0.5 and now we can clearly see the result which is 0.76. This information give us the idea that there is differences between those two groups of people, it is small difference but still exist. We can also see that the mean of the first group which is the "white people", it is a little bit lower than the second group "BMAE" people. This is the average age that is 36.18 for the white people, and 36.74 for the people from other ethnics. After that have been say, we can conclude that the difference between those groups exist. It seems that the infected white people are a little bit younger than the people from the other ethnics.

Through this simple test, I have managed to find the little difference that those two groups has between and that the infected white people are a little be younger. When we look at the raw data, it is hard to notice any difference but thanks to this test, we can always find this small different between any data.

Georgi Karachorov