



Julia Pietrocini

Student ID 2009998

Data Analysis and Visualisation 9ACCB4002)
Lecturer: Anupam Mazumbar



UNIVERSITY OF WALES
Trinity Saint David
London

Summary:

- What is the T-test?
- When can use the T-test?
- When can use the T-test?
- What this population?
- How to calculate the T-test?
- There are 50 white and BAME patients....
- Step by step... The last part!
- Results.
- Reference.

INTRUDACTION

In this second assessment of Data Analysis and visualisation, it's will report on the impact of covid 19 comparing white people with Bame people, based on age.

Taking into account the data provided, it will be possible to analyse and explain the processes that will guide the conduct of the T test.



What is the T-test?

Let's start with what a T-test is. The t-test is one of the most common statistical tests used to compare and relate two groups of subjects to each other. It allows you to check whether the mean value of a distribution deviates significantly from a certain preset value.

It is important to understand the target population being studied, so that you can understand who or what the data refer to. If you have not clearly defined who or what you want in your population, you may end up with data that is not useful to you. A statistical population is any group of individuals who are the subject of a study, which means that almost anything can constitute a population as long as the individuals can be grouped by a common characteristic, or sometimes by two common characteristics.

When can use the T-test?

For example, the T test can be used to assess whether there is a difference in average waiting times in emergency rooms between two different hospitals. You can also use this test to find out whether people who play a particular sport have, on average, a higher percentage of lean mass than those who do not. You can also use this test to find out whether there is a relationship between the average score on a mathematics exam and gender among students enrolled in a particular university.

The sample t-test determines whether the mean of the test group is different from the population mean. The data for a random t-test are normally distributed. The population samples should be random, and the sample size should be less than 30. The calculation for a random t-test gives the value. This value is used in a normal t table taking into account the degrees of freedom. The degrees of freedom are the number of the experimental group minus 1.

What this population?



The meaning of population on this assignment derives from individuality between the terms that were taken into consideration in a statistical way. The discussion about what population simply were taken under the scope beneath of the entire population

To get answers for this assignment one of the below 3 options should be chosen

T test for 2 sample assuming equal variances

T test for paired two sample means

T test for sample unequal variances

To create an analysis taken from above, there should be a method that will use a run test by applying different samples and formulas to gather information via Microsoft excel, and by doing a step to step analytical procedure on the formulas used.



Download from
Dreamstime.com

This watermarked comp image is for previewing purposes only.

ID 97171869

Bakhtiar Zein | Dreamstime.com

How to calculate the T-test?

1 Determine the null hypothesis. This should assume that there is no significant difference between the experimental mean and the population mean.

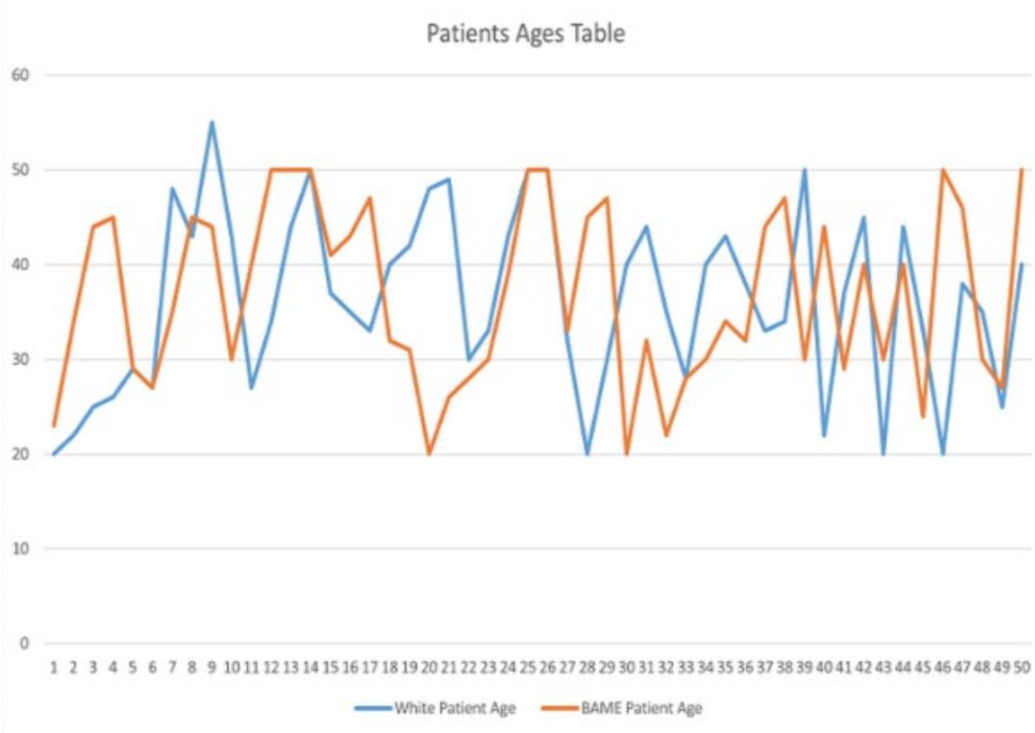
2 Calculate the mean of the experimental group. Calculate the standard deviation of the experimental group.

3 Subtract the experimental group ROM population mean. Divide by the standard deviation of the experimental group and multiply by the square root of the number in the experimental group.

4 Take this figure which is at the value. Compare this t-value with standard t-tables. A p-value less than 0.05 would be considered statistically significant.

-

There are 50 white and BAME patients, all aged between 20 and 50 years.



Patient No	White Patier	BAME	Patient No	White Patier	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

Step by step...

There are 50 white and bame patients, all aged between 20 and 50 years. The purpose of this report is to determine the infection between these two populations.

The process of starting a T-test consists first of all of dividing the data into two columns and assigning the right name and letter. The main column is the one indicating the number of patients. It will help to identify how many patients have been taken into consideration by dividing them into white and BAME groups.

The first group of whites is in the second column. In the third column the Bame taken into consideration.

The average values of both the white (X) and the Bame (Y) groups have to be established. The values are obtained by summing the whites with the Bame. Adding each number in the corresponding column. Once the calculations are complete, calculate the total per column and divide it by the patients.

The last part!

Once this step is done, start calculating the value of X AND Y which will come out by doing (X-x and Y-y), losing the values of each column. These values will be reported on a new color naming it XY.

After these steps, you have to find out the value of X and Y by following the column lines.

Calculated, the sum of XY is used to estimate the covariance value of xy (with x y) dividing it by the sum of the patients.

Calculating the power of X and Y, always following the column line. Using this formula we can estimate the value of the variance

$$\text{var } x = \frac{1}{N-1} \cdot \Sigma x^2 ; \text{ var } y = \frac{1}{N-1} \cdot \Sigma y^2$$

The last steps are to calculate the standard deviation of X and Y following two simple formulas $SDx = \sqrt{\text{var } x}$; $SDy = \sqrt{\text{var } y}$.

To obtain the variance of X we must calculate with the square root the value of X, same step for Y.

Obtaining all these values we can calculate the value of r that determines the correlation coefficient between X and Y.

After all the steps done before, observe the results.

	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	

Reference:

DREAMSTIME. (n.d.). Retrieved from <https://www.dreamstime.com/stock-illustration-population-growth-shows-family-reproduction-expecting-representing-newborn-image44556884>

EDUCBA. (2020). Retrieved from <https://www.educba.com/t-test-formula/>

Investopedia. (n.d.). Retrieved from <https://www.investopedia.com/terms/p/population.asp>

investopedia. (2020, march 22). Retrieved from <https://www.investopedia.com/terms/t/t-test.asp>

skuola.net. (n.d.). *skuola.net*. Retrieved from https://duckduckgo.com/?q=statistica&t=osx&iar=images&iax=images&ia=images&iai=https%3A%2F%2Fcdn.skuola.net%2Fnews_foto%2F2018%2Fstatistica-introduzione.jpg