

Data Analysis and Visualization

ACCB4002

Assignment 2

Student: Cristian Bitca(2013024)

Lecturer: Anupam Mazumdar

Table of Contents

Description of dataset	3
Hypotheses	4
Result of t Test with differit signifacance value	5
1. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.1$	5
2. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.01$	6
3. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.02$	7
References	8

Description of dataset

In a recent BBC (UK) news, it has been identified that the infection rate of Covid – 19 is varied between white people and BAME in the UK. In order to identify the ratio and the truthfulness of the findings, 50 different white and BAME patients have been selected. All the patients age is between 20 and 50 years. Based on the data in the table below, you are required to analyse the ratio of infection rate and truthfulness of the findings.

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

Figure 1 Dataset

The dataset above represents the proportion of the disease rate and the honesty found between 50 white patients and 50 BAME. All patients are between the ages of 20-50. This table was developed to investigate the rate of coronavirus contamination between white patients and BAME patines in the United Kingdom.

Hypotheses

The hypothesis is a model built only in natural language or with the help of mathematical language that tries to capture qualitatively or quantitatively the development of an event, to predict the existence of an object, property, or process, in reality, or conceptual space.

In this dataset, there are 50 white patients and 50 BAME patients we required to analyze the ratio of infection rate and truthfulness of the finding. Whether these 50 white patients and 50 Bame patients exist or do not exist between these two groups of patients scores differ to a statistical significance degree.

In statistical hypothesis testing, we compare two of them, named null hypothesis and alternative hypothesis.

In statistical hypothesis testing, we compare two of them, named null hypothesis and alternative hypothesis. consistent with Krantzler, John H. in his book, the null hypothesis is that interested in demonstrating the truth of some hypothesis of interest: that some relationship exists between two variables; that two groups differ in some important way. This hypothesis can take different forms, depending on the facts of the analyzed relation its named Alternative Hypothesis.

Null Hypothesis is noted with H_0 .

Alternative Hypothesis it is noted with H_1 .

The t Test is one of the most commonly used statistical tests. Its primary purpose is to determine whether the means of two groups of scores differ to a statistically significant degree.

We have 3 types of t Test:

- T Test Paired Two Sample for Mean
- T Test Two-Sample Assuming Equal Variance
- T Test Two-Sample Assuming Unequal Variance

Tips to remember:

1. If your dataset is very small and you can easily identify their mean values then go for an **equal t-test**.
2. If your data set is not relatively small and difficult to identify means the go for an **unequal t-test**.

Rules:

1. If your P values is smaller than the alpha value then you must reject your null hypothesis.
2. If your P values is greater than the alpha value then you have to accept the null hypothesis.

Possibility if we use equal t test:

H_0 = There is no significant differences ratio of infection between white patients and BAME patients.

H_1 = There is significant differences ratio of infection between white patients and BAME patients.

Possibility if we use unequal t test:

H_0 = There is significant differences ratio of infection between white patients and BAME patients.

H_1 = There is no significant differences ratio of infection between white patients and BAME patients.

Result of t Test with differit signifacance value

1. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.1$

	<i>Variable 1</i>	<i>Variable 2</i>
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: Two-Sample Assuming Unequal Variances with significance value $p=0.1$

In the case where the value of p is 0.1, I performed the test and found out the following result.

I chose the unequal t test option for that:

H_0 = There is significant differences ratio of infection between white patients and BAME patients.

H_1 = There is no significant differences ratio of infection between white patients and BAME patients.

- a. T Stat = -0.298879822 and t Critical one-tail = 1.290249904
- b. P two value is 0.76566461 rounded and significance value is 0.1

Result : We will accept null hypothesis

2. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.01$

t-Test: Two-Sample Assuming Unequal Variances		
	<i>Variable 1</i>	<i>Variable 2</i>
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: Two-Sample Assuming Unequal Variances with significance value $p=0.01$

In the case where the value of p is 0.01, I performed the test and found out the following result.

I chose the unequal t test option for that:

H_0 = There is significant differences ratio of infection between white patients and BAME patients.

H_1 = There is no significant differences ratio of infection between white patients and BAME patients.

- a. T Stat = -0.298879822 and t Critical one-tail = 0.382832305
- b. P two value is 0.76566461 rounded and significance value is 0.1

Result : We will accept null hypothesis

3. t-Test: Two-Sample Assuming Unequal Variances with significance value $p = 0.02$

	<i>Variable 1</i>	<i>Variable 2</i>
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	2.081444921	
P(T<=t) two-tail	0.76566461	
t Critical two-tail	2.36500241	

Figure 4 t-Test: Two-Sample Assuming Unequal Variances with significance value $p=0.02$

In the case where the value of p is 0.01, I performed the test and found out the following result.

I chose the unequal t test option for that:

H_0 = There is significant differences ratio of infection between white patients and BAME patients.

H_1 = There is no significant differences ratio of infection between white patients and BAME patients.

- a. T Stat = -0.298879822 and t Critical one-tail = 2.081444921
- b. P two value is 0.76566461 rounded and significance value is 0.02

Result : We will accept null hypothesis

References

Statistics for the terrified.(Kranzler, John H.) available at -> <https://www-vlebooks-com.ezproxy.uwtsd.ac.uk/Vleweb/Product/Index/990670?page=0>