e-ISSN: 2456-3463
International Journal of Innovations in Engineering and Science, Vol. 2, No.2, 2017
www.ijies.net

Application of t-test to analyze the small sample of Statistical Research

N.M. Bhondge[1], L.H.Sarode[2]

[(1)]Dept. of Mathematics, Priyadarshini Indira Gandhi College of Engineering, Nagpur
Email: nitinbhondge@gmail.com

[(2)]Dept. Of psychitric, Asharam College of Nursing, Kamptee, Nagpur
Email : leenasarode6@gmail.com

Abstract -*In this paper we discuss the significance of t-test in small sample of statistics to analyze the data. Here we approach many application of t-test in statistics and research. This paper is aimed at introducing hypothesis testing, focusing on the paired t-test. It will explain how the paired t-test is applied to statistical analyses.*

Keywords: t-test, p-value, significance, hypothesis testing.

INTRODUCTION

Hypothesis testing is used to make an inference about a population that's under study. The inference is based on the parameter(s) for the statistic, usually the sample mean and standard deviation. Suppose it is believed that the mean of a population is zero, the first step in hypothesis testing is to state the null hypothesis (H_0) and an alternative hypothesis (H_a). The null hypothesis is the assumption that the mean will be equal to zero. The alternative hypothesis is the assumption that the mean will be either greater than zero, less than zero, or simply, not equal to zero. When the alternative hypothesis! states that the mean is less than zero, the test is called a left-tailed test. It is right tailed! when H a states that the mean is greater than zero. The test is called two-tailed when H_a states that the mean is not equal to zero. The second step in hypothesis testing is to calculate a test statistic. Based on the value of the test statistic, you will find the p-value from the table of z-distributions, which is based upon the normal distribution known as the bell curve.

The determined p-value is compared to the level of significance, α (alpha). The level of significance is the probability of making a Type I error. A Type I error occurs when H_0 is rejected when it is true. If H_0 is not rejected when it is actually false, the error is called as a Type II error. The alpha level can be computed as one minus the confidence level. The confidence level is equivalent to the area under the curve. Common values for the confidence level are .90, .95, and .99. Of the three, the most! Commonly used is the confidence level at .95. This suggests a 95% confidence that the null hypothesis is true. Thus, the level of significance that is commonly used is .05(1 -.95). Thus, α set at .05 suggests that there is a 5% chance of making a Type I error. If the p-value is less than or equal to α , the null hypothesis should be rejected. If the p-value is greater than α , H_0 should not be rejected. The last step is to explain the results of the test and what it concludes about the analysis.

The terms “significance level” or “level of significance” refer to the likelihood that the random sample you choose (for example, test scores) is not representative of the population. The lower the significance level,the more confident you can be in replicating your results. Significance levels most commonly used in educational research are the .05 and .01 levels. If it helps, think of .05 as another way of saying 95/100 times that you sample from the population, you will get this result. Similarly, .01 suggests that 99/100 times that you sample from the population, you will get the same result. These numbers and signs (more on that later) come from Significance Testing, which begins with the Null Hypothesis.

The paired t-test is a type of hypothesis testing that is used when two sets of data are being observed. The data in a paired t-test are dependent, because each value in the first sample is paired with a value in the second sample. The parameter used to make the inference is the difference of the means of both data sets. Similar to our previous hypothesis test example, the null hypothesis states that the

difference of the means is equal to zero ($\mu_1 - \mu_2 = 0$) . This can also be understood as the means are equal. The alternative hypothesis can be the mean of the first sample is greater than the mean of the second ($\mu_1 - \mu_2 > 0$) , the mean of the first sample is less than the mean of the second sample ($\mu_1 - \mu_2 < 0$) , or the means are not equal with no greater than or less than comparison ($\mu_1 - \mu_2 \# 0$) .

The Student's t-test is the test statistic used in a paired t-test. The Student's t is a distribution discovered by a statistician, W.S. Gosset. Because he worked for a company that did not approve of its employees publishing their research, when publishing his work he used the name Student. When a population is too large for testing, samples are taken from the population and used in a test. Sampling is expensive and the t-distribution is ideal when the sample is small (i.e. < 30). The t-distribution was created for use when σ is unknown. The student t-distribution is similar to normal distribution in that it is bell shaped and symmetrical. Its shape depends on the degrees of freedom, which is one less than the sample size (n -1). As degrees of freedom increase, the t-distribution approaches normal distribution. Gosset developed the table of t-distributions to find the p-value that corresponds to the test statistic and the degrees of freedom. The values in the Student t-table are calculated by formulas developed by Gosset. This falls under another! subject called mathematical statistics, and is beyond the scope of this scholarly note.
Formula to calculate the t-test statistic: $d=x_1-x_2$ difference

Here we introduce the case study on caregivers of depressive patients with distribution of caregivers of depressive patients with regards to their demographic variables. A convenient sample of 60 subjects was drawn from the study population, who were selected from selected hospitals of city. The data obtained to describe the sample characteristics including age, gender, type of family, education, occupation and source of information respectively.

Demographic Variables	**Frequency**	**Percentage (%)**
Age(yrs)		
18-26	6	10.0
27-35	10	16.7
36-44	25	41.7
45-53	13	21.7
54-62	6	10.0
Gender		
Male	21	35.0
Female	39	65.0
Type of family		
Nuclear	20	33.3
Joint	40	66.7
Education		
Primary	15	25.0
Secondary	21	35.0
Higher Secondary	18	30.0
Graduate	5	8.3
PG	1	1.7
Other	0	0.00
Occupation		
Housewife	28	46.7
Private Employee	16	26.7
Govt. Employee	7	11.7
Others	9	15.0
Source of information		
Mass Media	5	8.3
Health Personnel	30	50.0
Friends	7	11.7
Relatives	12	20.0
Others	6	10.0

This section deals with the assessment of pre test knowledge regarding depression among care givers of depressive patients in selected hospitals of city.

Assessment of post test knowledge regarding depression among care givers of depressive patients.

Level of knowledge score	**Score range**	**Percentage score**	**Post Test**	
			Frequency	**Percentage**
Poor	1-5	0-19%	0	0.00
Average	6-10	20-39%	0	0.00
Good	11-15	40-59%	1	1.67
Very Good	16-20	60-79%	15	25.00
Excellent	21-25	≥80%	44	73.33
Minimum score	14			
Maximum score	25			
Mean score	20.61±2.40			
Mean %	82.46±9.60			

Now here we deals with the effectiveness of planned teaching programme on knowledge regarding depression among care givers of depressive patients in selected hospitals of city. The hypothesis is tested statistically with area wise distribution of pre test and post test mean and standard deviation mean difference in pre and post test knowledge score. The levels of knowledge during the pre test and post test are compared to prove the

effectiveness of planned teaching programme. Significance of difference at 5% level of significance is tested with student's paired't' test and tabulated 't' value is compared with calculated 't' value. Also the calculated 'p' values are compared with acceptable 'p' value i.e. 0.05.

Overall	Mean	SD	SE	Mean Percentage	t-value	p-value
Pre Test	8.23	3.14	0.40	32.93	22.03	0.000 S,p<0.05
Post Test	20.61	2.40	0.31	82.46		

Association of knowledge score in relation to demographic variables.

Age (yrs)	No. of caregivers of depressive patients	Mean post knowledge score	F-value	p-value
18-26	6	21.16±1.47	0.86	0.49 NS ,p>0.05
27-35	10	20.80±2.97		
36-44	25	21.04±2.49		
45-53	13	19.69±2.21		
54-62	6	20±2.09		

This table shows the association of post test knowledge scores with the age of caregivers of depressive patients. The tabulated 'F' values was 2.52(DF=4, 59) which is higher than the calculated 'F' i.e. 0.86 at 5% level of significance. Also the calculated 'p'=0.49 which was much higher than the acceptable level of significance i.e. 'p'=0.05. Hence it is interpreted that the age of the care givers of depressive patients is not associated with their post test knowledge scores.

Gender	No. of caregivers of depressive patients	Mean post knowledge score	t-value	p-value
Male	21	20.52±1.77	0.21	0.82 NS,p>0.05
Female	39	20.66±2.69		

This table shows the association of post test knowledge scores with the gender of caregivers of depressive patients. The tabulated 't' values was 1.98(df=58) which is higher than the calculated 't' i.e. 0.21 at 5% level of significance. Also the calculated 'p'=0.82 which was much higher than the acceptable level of significance i.e. 'p'=0.05. Hence it is interpreted that the gender of the caregivers of depressive patients is not associated with their post test knowledge scores.

Type of family	No. of caregivers of depressive patients	Mean post knowledge score	t-value	p-value
Nuclear	20	20.30±2.88	0.71	0.47 NS,p>0.05
Joint	40	20.77±2.14		

This table shows the association of post test knowledge scores with the type of family of caregivers of depressive patients. The tabulated 't' values was 1.98(df=58) which is higher than the calculated 't' i.e. 0.71 at 5% level of significance. Also the calculated 'p'=0.47 which was much higher than the acceptable level of significance i.e. 'p'=0.05. Hence it is interpreted that the type of family of the caregivers of depressive patients is not associated with their post test knowledge scores.

This table shows the association of post test knowledge scores with the educational status of caregivers of depressive patients. The tabulated 'F' values was 2.52(df=4,59) which is much less than the calculated 'F' i.e. 9.71 at 5% level of significance. Also the calculated 'p'=0.000 which was much less than the acceptable level of significance i.e. 'p'=0.05. Hence it is interpreted that the educational status of the care givers of depressive patients is statistically associated with their post test knowledge scores.

Occupation	No. of caregivers of depressive patients	Mean post knowledge score	F-value	p-value
Housewife	28	20.10±2.57	2.42	0.075 NS,p>0.05
Private Employee	16	20.75±1.87		
Govt. Employee	7	22.71±1.25		
Others	9	20.33±2.73		

This table shows the association of post test knowledge scores with the occupational status of caregivers of depressive patients. The tabulated 'F' values was 2.76(df=3,59) which is much higher than the calculated 'F' i.e. 2.42 at 5% level of significance. Also the calculated 'p'=0.075 which was much higher than the acceptable level of significance i.e. 'p'=0.05. Hence it is interpreted that the occupational status of the care givers of depressive patients is not associated with their post test knowledge scores.

Source of information	No. of caregivers of depressive patients	Mean post knowledge score	F-value	p-value
Mass Media	5	22±1.22	0.55	0.70 NS ,p>0.05
Health Personnel	30	20.30±2.56		
Friends	7	20.57±1.71		
Relatives	12	20.75±2.73		
Others	6	20.83±2.48		

This table shows the association of post test knowledge scores with the source of information about depression of caregivers of depressive patients. The tabulated 'F' values was 2.52(df=4,59) which is much higher than the calculated 'F' i.e. 0.55 at 5% level of significance. Also the calculated 'p'=0.70 which was much higher than the acceptable level of significance i.e.'p'=0.05. Hence it is interpreted that the source of information of care givers of depressive patients is not associated with their post test knowledge scores.

CONCLUSION

After the detailed analysis, this study leads to the conclusion that the care givers do not have 100% knowledge regarding depression. There was a significant increase in the knowledge of subjects after the introduction of planned teaching. To find the effectiveness of planned teaching's' test was applied and t value was calculated, post test score was significantly higher at 0.05 level than that of pre test score. Thus it was concluded that planned teaching on depression was found effective as a teaching strategy.

REFERENCES

[1] Stephanie D. Wilkerson, Mathematics 'Application of the Paired t-test' XULAneXUS: Xavier University of Louisiana's Undergraduate Research Journal. Scholarly Note. Vol. 5, No. 1, April 2008

[2] Sharma Suresh K. Nursing research and statistics. 1st edition 2011, Elsevier publisher, printed at Harayana, page number: 70-73

[3] World Health Organization. The world health report 2001 – Mental Health: New Understanding, New Hope; 2001 [Retrieved 2008-10-19].

[4] Joginder Kaur Lecturer (Mathematics), Punjab Technical University, Punjab, India 'Techniques Used in Hypothesis Testing in Research Methodology – A Review'International Journal of Science and Research (IJSR)ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

[5] J.C.F. de Winter Delft University of Technology 'Using the Student's t -test with extremely small sample sizes' Practical Assessment, Research & Evaluation. Volume 18, Number 10, August 2013, ISSN 1531-7714