Vol.9.Issue.4.2021 (Oct-Dec)
©KY PUBLICATIONS

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

The Validity of t-test and Z-test for Small One sample and Small Two Sample test

Dr.Ramnath Takiar
Scientist G – (Retired)
National Centre for Disease Informatics and Research (NCDIR),
Indian Council of Medical Research (1978-2013)
Bangalore – 562110, Karnataka, India.
DOI:10.33329/bomsr.9.4.42


Dr.Ramnath Takiar

ABSTRACT

The present study was carefully designed to evaluate the performance of t-test as compared to Z-test in testing the significant or non-significant differences between two sample means. The sources of data for the study came by generating four Normal populations (Population A, B, C and D) and then drawing 30 samples each form those populations. Overall, the study covers 14400 comparisons to test for significant differences and 18240 comparisons for non-significant differences between means.

It is surprising to note that at α = 5%, t-test was able to pick up only 29.3% of the expected significant differences between the sample means of Population C and D, which is quite low. In case of Population A and B, the validity of the test was observed to be relatively better and it was 49.6%.

In view of low validity observed in the case of α = 5%, the validity was further explored at the higher levels of α namely 10%, 15% and 20%. With the rise in α levels, the validity was observed to be increasing. For the Population C and D, at α = 20% , the validity of t-test rose to 54.3% and for the Population A and B, the validity rose to 76.1%. This suggests that for testing the significant or non-significant differences between the means, especially for small samples, the α level can be raised from 5% to 20% so that more valid mean comparisons by t-test can be obtained.

In view of Z-test performing better as compared to t-test in picking up the significant differences, correctly, and not lagging behind much in picking up the non-significant differences between two sample means, suggests that Z-test can be used even for small sample sizes in place of hitherto used t-test.

Keywords: t-test, Z-test, Simulation, Normal samples, Validity, 5% α level, 20% α level.

INTRODUCTION

In medical research, t-test is widely used to decide whether two sample means obtained in connection of some research study or survey are comparable or not? According to W.S. Gosset (1908) if the sample size is small say less than 30 then the Z statistics namely Z = $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ or Z = $\frac{\bar{x}-np}{\sqrt{npq}}$ do not behave like normal distribution and as such the "normal test", also known as Z test, cannot be applied. In most of the applications, the sample mean is taken as the estimate of the population mean. Similarly, when σ is unknown, we can obtain an estimate S of σ from the sample that give us the estimate of mean. According to W.S. Gosset, if the sample is of size n, then the statistic t = $\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$ follows t-distribution with n-1 degree of freedom. Subsequently, t-test application was also extended for testing the difference between two sample means (Snedecor and Cochran 1967, Gupta and Kapoor, 2001, Gupta 2012).

In case of one sample, the interest lies in determining whether the sample comes with a specified mean or not? Similarly, in case of two independent samples, the interest lies in comparison and in determination whether two samples have comparable means or not? Prior to discovery of t-test, in both the situations, the Normal test was in use. With the emergence of t-test, especially for small samples, t-test was advocated. However, it was never determined what is the loss, if we continue to use the Normal test even for small samples or how much we gain in validity for going t-test in place of Normal test. Therefore, a study was designed to answer the questions raised with the following objectives:

1. To assess the validity of the t-test as compared to Z-test, for a set of sample means, in demonstrating that those means are not significantly different from the population mean μ when σ is unknown.
2. To assess the validity of t-test in comparison to Z-test, for two sets of sample means, in picking up the significant differences or non-significant differences between them.

MATERIAL AND METHODS

SOURCES OF DATA

For the t-test applications whether it is a single mean testing or testing of difference between two sample means, we need samples of data to proceed. Further, it is essential that those samples are drawn from the Normal populations. Using the simulation technique, the required normal populations as well as the samples are generated and used as a source of data. Thus, for the study purposes, the sources of data come through the simulation technique.

SIMULATIONS OF NORMAL POPULATIONS

For generation of Normal samples, the function key, "Random Number Generation" provided in StatPlus 7.6.5 was used. This function key prompt you to provide the name of the distribution in which you are interested. From the list of distribution, select "Normal". Further, you will be prompted to provide the information on: Number of new variables, Random Number Count, Mean and Standard Deviation. On filling the appropriate information, you will get a set of Normal values. For example, if you fill, 25, 100, 55.5, 16.05, in above prompts, then you will get 25 random samples of size 100, with the Population mean = 55.5 and SD = 16.05. The sample mean and SD may show some variation but for all practical purposes they remain as a sample drawn from a normal population with a specified mean and SD. For the study purposes, two typical random samples of size 100 were so chosen that when they are mixed, give rise to a normal population with skewness close to 0 and Kurtosis close to 3.0. Proceeding in a similar way, the following four sets of Normal values were generated and termed as the Population A, Population B, Population C and Population D and shown in Table 1.

It is ensured that the population means of A and B are significantly different. Similarly, the population means of C and D are ensured to be significantly different. Going by the characteristics provided, there is no doubt left that each of the population is very close to Normal population especially when judged by the values of Skewness and Kurtosis, expected to be 0 and 3.0, respectively for a Normal Population. Based on the significant differences found between (A and B) and (C and D), it is safe to assume that they are different Normal Populations of size 200, each.

Thus, the basic details of the four Normal populations, generated, are provided below:

Table 1: Some Details of Populations, Selected

Population	A	B	C	D
Population size	200	200	200	200
Mean	55.5	44.21	65.77	76.14
SD	16.05	11.73	16.824	17.99
Minimum	16.42	9.17	14.42	27.2
Maximum	97.66	71.59	106.63	126.5
Skewness	0.02	-0.11	0.06	0
Kurtosis	2.9	3.0	2.97	3.03
Z Value	8.03		5.95	
P-value	< 0.001		< 0.001	

SAMPLE SELECTION

The scheme of sample selection according to Population, Number of Samples and the Sample size is shown in Table 2.

As per the Table 2, for each population, 30 samples were simulated in total with varying sample sizes of 9,13 and 20. The random samples of given size were generated using the function key of "Random Sample" provided with StatPlus 7.6.5. This function key allows you to select a random sample of the chosen size from the desired set of numbers provided. In the present context, it is the Population A, B, C and D.

SCHEME OF COMPARISONS OF THE SAMPLE MEANS AND POPULATION MEAN

From the Population A, 10 Random samples each of size 9, 13 and 20 were drawn. Each sample mean allows us to be compared with the Population Mean of A, thus allowing for 30 t-tests and equal number of Z-test. Proceeding, in a similar way, for all the four Populations (A, B, C and D), 120 t-tests or Z-tests can be carried out.

Table 2: Scheme of Sample Selection according to Population, Number of Samples and the Sample size

Population	Sample Size	Number of Samples Drawn
A	9	10
	13	10
	20	10
B	9	10
	13	10
	20	10
C	9	10
	13	10
	20	10
D	9	10
	13	10
	20	10
Total	Pooled	120

SCHEME OF COMPARISONS BETWEEN THE MEANS OF DIFFERENT POPULATIONS

The number of possible mean comparisons according to different sample size and Population is shown in the Table 3.

Table 3: Number of Mean Comparisons according to Different Sample size and the Population

Population A	Population B	Number of Mean comparisons	Population C	Population D	Number of Mean comparisons
Sample size	Sample size		Sample size	Sample size	
9	9	100	9	9	100
	13	100		13	100
	20	100		20	100
13	9	100	13	9	100
	13	100		13	100
	20	100		20	100
20	9	100	20	9	100
	13	100		13	100
	20	100		20	100
Total		900	Total		900

Note: There are 10 samples for each sample size

The comparison of set of 10 sample means of Population A, for a given sample size, with those of set of sample means of varying sample size of Population B, allow us for 300 comparisons, leading to a total of 900 comparisons between the means of Population A and B. Similarly, a total of 900 mean comparisons can be carried out for the Population of C and D. Thus, in total, for all the four populations, 1800 comparisons between the means can be carried out by the t-test and the Z-test.

SCHEME OF COMPARISONS BETWEEN THE MEANS OF SAME POPULATIONS

The number of mean comparisons made according to various combinations of sample sizes within a Population and by different Populations are shown in Table 4. Allowing the comparisons between all the possible pairs of mean, for a given Population, as shown in the table below, we expect to have comparisons of 570 means. Thus, for all the four populations, a total of 2280 mean comparisons can be made by the t-test as well as by the Z-test.

Table 4: Number of Mean Comparisons according to Various Combinations of Sample size Within a Population and by Different Populations

Population	Sample size	Sample size			Total
		9	13	20	
A	9	90	100	100	290
	13	-	90	100	190
	20	-	-	90	90
B	9	90	100	100	290
	13	-	90	100	190
	20	-	-	90	90
C	9	90	100	100	290
	13	-	90	100	190
	20	-	-	90	90
D	9	90	100	100	290
	13	-	90	100	190
	20	-	-	90	90
Total		360	760	1160	2280

VALIDITY OF T-TEST AND Z TEST FOR THE COMPARISONS OF THE SAMPLE MEANS WITH THAT OF THE POPULATION MEAN

In this case, the Null Hypothesis formulated will be that "The sample mean is not different from the specified Population Mean". As all the sample means are tested against their own known Population Mean, it is logical that we should be failing to reject the Null Hypothesis. Thus, a higher percentage (≥ 80%) of non-significant tests will suggest a higher validity of the t-test. Similarly, a lower percentage of non-significant tests (say, below 70%) will suggest a lower validity of the t-test. In a similar way, the validity of the Z-test is also assessed.

VALIDITY OF T-TEST AND Z-TEST FOR COMPARISONS OF THE SAMPLE MEANS WHEN DRAWN FROM THE SAME POPULATION

In this case, again, the Null Hypothesis formulated will be that "The sample means are not significantly different from each other". As the sample means that we are comparing, are known to come from the same Normal Population, we are expected not to reject the Null Hypothesis. Thus, A higher percentage of non-significant differences will suggest that the t-test or Z-test is successful in picking up the desired non-significant differences between the sample means, correctly. A low percentage will indicate the lower validity of the t-test or that of Z-test.

VALIDITY OF T-TEST AND Z-TEST FOR COMPARISONS BETWEEN THE MEANS OF THE SAMPLES DRAWN FROM TWO DIFFERENT NORMAL POPULATIONS

In this case, as the sample means that we are comparing, are known to come from the two different Normal Populations, hence expected that the Null Hypothesis "The sample means are not significantly different from each other" will be rejected and the alternative Hypothesis that the sample means are significantly different from each other is accepted. A higher percentage of significant differences between the means suggests the higher validity of the t-test or Z-test while a lower percentage say below 70% will suggest that the validity of the test is below 70%.

ANALYSIS OF THE DATA

According to the objectives of the study, the use of the t-test was necessary to test the significant or otherwise non-significant differences between the means of the two samples irrespective of when samples are drawn from the same population or from the two different populations. A small Excel program, developed by me to test the significance differences between two samples by t-test, was utilized. The function key available on Excel 2019, was utilized to arrive at the probability of the t-statistic calculated.

EXCEL PROGRAM AND USE OF MACRO

A small macro was developed to find out the t-statistic and corresponding probability while attempting to assess the significance differences between two sample means, for various pairs of samples, at a time. It is worth mentioning here that overall, 14400 comparisons were made to test for significant differences and 18240 comparisons for non-significant difference for the study purposes which became possible only by the use of an appropriate Excel program and a macro (Excel 2019). The results obtained by t-test and Z-test are also compared to see which test is better in picking up either the significant differences or the non-significant differences.

RESULTS

To give an idea to the readers about the simulated random samples, two typical random samples, out of 10, of size 9, 13, 20, each, drawn from each of the Population A, B, C and D, are shown in Table 5, 6 and 7.

Table 5: Typical two Random Samples from Each Population of Size 9

	Population A		Population B		Population C		Population D	
	S1	S2	S1	S2	S1	S2	S1	S2
1	55.43	56.15	48.49	26.06	61.76	76.65	78.4	82.4
2	76.57	67.72	34.03	48.52	90.1	63.05	47.5	75.7

3	17.83	48.07	64.43	53.7	70.02	100.7	74	77.3
4	42.11	72.89	40.03	42	68.87	66.9	99.8	84.7
5	40.84	53.13	46.83	44.96	41.44	70.02	74.1	91.4
6	48.07	31.24	47.81	18.78	80.3	68.64	57.2	84.8
7	60.09	33.15	49.75	55.89	54.71	80.71	78.9	27.8
8	67.55	71.98	66.05	46.78	57.37	69.31	60.9	58.8
9	60.09	54.39	48.99	40.34	78.31	69.84	88.9	81.1
Mean	52.06	54.30	49.60	41.89	66.99	73.98	73.30	73.78
SD	17.297	15.252	10.229	12.247	14.898	11.287	16.135	19.473

Table 6: Typical two Random Samples from Each Population of Size 13

	Population A		Population B		Population C		Population D	
	S1	S2	S1	S2	S1	S2	S1	S2
1	75.06	53.68	52.85	29.71	96.6	42.9	80.2	76.6
2	33.9	60.04	36.5	47.82	84.5	48.9	96.8	82.2
3	51.92	49.82	43.89	49.18	49.6	14.4	84.4	108.6
4	20.25	71.52	29.08	46.82	68.2	55.5	77	71.5
5	79.03	57.68	41.11	31.39	80.7	49.8	77	76.9
6	38.04	66.79	53.63	34.03	76.7	29.1	81.2	83.4
7	31.16	33.9	71.31	32.14	93.1	61.2	85.5	83
8	51.05	66.39	29.55	24.66	65.7	84.5	72.2	56.3
9	49.6	59.56	42.11	34.16	76.5	45	57.2	71
10	52.71	40.53	39.76	63.98	52.8	52.8	46.3	61.7
11	60.02	46.67	47.78	49.56	83.3	88.5	107.4	72.2
12	49.03	46.27	33.87	40.87	79.4	68.1	82.3	85.2
13	80.35	39.64	40.35	62.86	68.9	65.6	77.7	112.9
Mean	51.70	53.27	43.21	42.09	75.08	54.33	78.86	80.12
SD	18.507	11.622	11.39	12.489	13.899	20.298	15.308	16.032

Table 7: Typical two Random Samples from each Population of Size of 20

	Population A		Population B		Population C		Population D	
	S1	S2	S1	S2	S1	S2	S1	S2
1	49.03	42.11	41.56	56.81	60.1	72.4	75.3	92
2	63.05	46.62	34.69	71.59	63.4	66.6	73.9	72.2
3	55.72	53.34	29.38	56.35	100	102.5	44.7	75.7
4	66.29	60.81	46.29	37.54	89.1	58	59.2	76.3
5	43.08	88.57	47.96	71.31	96.8	49.3	77.4	17.99

6	51.93	69.84	58.5	24.32	65.7	92.3	72.7	64.2
7	57.13	29.04	56.01	47.7	77.3	77	77.7	68.6
8	80.7	30.5	30.27	42	84.1	93.8	108.3	49
9	68.87	64.6	37.38	40.35	55.4	76.7	49.2	76.2
10	48.96	92.12	63.65	34.03	89.1	52.1	53.1	82.8
11	55.43	51.93	41.61	41.73	40.6	66.9	81.2	71.7
12	70.35	97.66	27.07	54.64	105.3	89	65.7	92.6
13	52.32	64.27	43.88	29.55	59	67.3	40.6	97.1
14	47.31	57.68	41.11	55.34	80.9	49.6	58.8	108.6
15	77.29	39.64	48.22	44.17	80.8	94	62.6	98.6
16	60.81	48.94	29.22	48.99	77.2	74.4	71	116.8
17	30.5	47.31	54.64	32.09	113.9	74	73.2	80.8
18	35.82	57.77	55.47	52.19	73.9	70.4	49	62
19	46.67	59.42	32.65	48.14	70.9	45.6	117.1	66.7
20	59.76	51.92	55.34	56.01	75.3	41.4	98.6	46.3
MEAN	56.05	57.70	43.75	47.24	77.94	70.67	70.47	75.81
SD	12.865	18.508	11.093	12.615	18.068	17.673	20.235	22.675

The details of Mean and SD, of each 10 samples, according to different sample sizes of Population A and B are provided in Table 8 and 9, respectively.

The details of Mean and SD, of each of the 10 samples, according to different sample sizes of Population C and D are provided in Table 10 and 11, respectively.

Thus, the set of Means and SDs provided in Table 8, 9, 10 and 11, actually formed the data for testing the validity of the t-test and Z-test.

Table 8: Details of Mean and SD According to Different Sample Sizes - Population A
(Mean = 55.5; SD = 16.05)

Sample size	Statistic	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
9	Mean	52.1	54.3	46.8	57.3	52.6	54.8	52.2	55.8	57.4	65.5
	SD	17.30	15.25	19.13	19.44	8.11	18.14	22.46	10.44	12.4	21.59
13	Mean	51.7	53.3	57.6	46.8	58.6	54	52.8	44.4	46.9	52.7
	SD	18.51	11.62	15.73	12.96	13.62	17.09	13.65	20.39	21.48	12.28
20	Mean	56.1	57.7	51.8	61.5	53.9	56.4	53	57.2	56.3	53.6
	SD	12.87	18.51	16.89	16.13	16.44	19.66	18.37	18.25	17.31	11.96

Table 9: Details of Mean and SD According to Different Sample Sizes - Population B
(Mean = 44.2; SD = 11.73)

Sample size	Statistic	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
9	Mean	49.6	41.9	40.6	44.7	45.3	47.6	44.7	44.4	42.6	43.4
	SD	10.23	12.25	12.69	10.86	12.08	9.76	4.41	13.07	14.53	10.44
13	Mean	43.2	42.1	43.9	43.4	36.8	41.6	40.8	40.3	44.6	44.2
	SD	11.39	12.49	11.35	11.02	12.09	11.14	14.45	11.87	6.5	15.15
20	Mean	43.7	47.2	39.5	47.8	42.3	43.4	41.9	44.1	47.1	44.6
	SD	11.09	12.62	12.69	10.9	13.23	9.79	13.57	12.16	14.99	8.8

Table 10: Details of Mean and SD According to Different Sample Sizes - Population C
(Mean = 65.8: SD = 16.82)

Sample size	Statistic	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
9	Mean	67.0	74.0	72.7	63.2	73.3	53.5	60.2	73.4	65.6	57.9
	SD	14.90	11.29	20.52	17.28	13.97	16.36	9.76	18.80	13.66	17.39
13	Mean	75.1	54.3	63.1	66.4	65.7	72.1	64.4	71.4	63.0	64.8
	SD	13.90	20.30	9.92	16.62	21.39	11.27	17.16	14.76	16.96	18.12
20	Mean	77.9	70.7	66.1	63.5	68.5	71.0	70.3	67.4	64.9	61.3
	SD	18.07	17.67	18.37	11.75	11.28	16.86	16.83	17.16	20.25	15.06

Table 11: Details of Mean and SD According to Different Sample Sizes - Population D
(Mean = 76.1; SD = 17.99)

Sample size	Statistic	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
9	Mean	73.3	73.8	82.0	77.0	78.2	77.4	79.8	75.6	74.2	82.7
	SD	16.14	19.47	20.68	8.95	9.68	24.99	22.64	17.89	21.93	10.08
15	Mean	78.9	80.1	70.5	75.3	77.7	76.7	71.7	79.2	81.3	77.6
	SD	15.31	16.03	18.57	18.79	21.45	17.95	15.85	14.04	18.01	21.54
20	Mean	70.5	75.8	73.1	75.2	80.7	76.0	74.1	78.2	80.5	71.1
	SD	20.24	22.68	15.27	20.75	15.75	15.52	18.25	15.29	17.10	15.39

The results of t-test and Z test to compare differences between different pairs of sample means, at different alpha levels when drawn from two different populations (A and B), are shown in Table 12. Over all, at 5% α level, when t-test was applied between different pairs of means, t-test was able to pick up only 49.6% the expected significant differences correctly, as against 54.8% of significant differences, picked up correctly by the Z test. On raising the α level to 10%, the significant differences picked up by t-test rose to 63.4%. At $\alpha =$ 15% and 20%, the correct percentage of significant differences rose to 71.2 % and 76.1% respectively. In general, Z-test was found to be picking up more significant differences than t-test.

Table 12: Results of t-test as Compared to Z test, for the difference between two means by different sample sizes and alpha levels – (Population A and B)

Sample	5%		10%		15%		20%	
	t-test	Z-test	t-test	Z-test	t-test	Z-test	t-test	Z-test
9A - 9B	26	38	44	54	55	62	64	69
9A - 13B	58	65	71	74	83	83	85	86
9A - 20B	50	55	61	64	70	72	74	79
13A - 9B	23	22	37	42	46	51	54	55
13A - 13B	52	54	62	62	69	69	72	72
13A - 20B	46	49	52	53	55	56	61	61
20A - 9B	37	46	62	66	73	77	81	85
20A - 13B	84	87	95	94	99	99	100	100
20A - 20B	70	77	87	90	91	92	94	94
Significant	446	493	571	599	641	661	685	701
Total test	900	900	900	900	900	900	900	900
%	49.6	54.8	63.4	66.6	71.2	73.4	76.1	77.9

9A - Sample size = 9 drawn from the Population A; 9B - Sample size = 9 drawn from the Population B

The results of t-test and Z test to compare the differences between paired sample means, at different α levels when drawn from two different populations (C and D), are shown in Table 13.

Table 13: Results of t-test as Compared to Z test, for the difference between two means by different sample sizes and a levels (5% - 20%)

Sample	5%		10%		15%		20%	
	t-test	Z-test	t-test	Z-test	t-test	Z-test	t-test	Z-test
9C - 9D	28	37	42	46	46	49	49	55
9C - 13D	34	39	41	44	47	49	50	53
9C - 20D	32	37	39	40	44	46	46	50
13C - 9D	24	35	38	47	49	57	58	62
13C - 13D	33	40	45	49	54	63	64	68
13C - 20D	30	35	42	45	52	56	56	61
20C - 9D	22	28	40	43	46	47	50	55
20C - 13D	31	33	38	46	55	58	62	63
20C - 20D	29	31	39	42	47	50	54	58
Significant	263	315	364	402	440	475	489	525
Total test	900	900	900	900	900	900	900	900
%	29.2	35.0	40.4	44.7	48.9	52.8	54.3	58.3

For the populations (C and D), at 5% α level, over all, t-test was able to pick up only 29.2% significant differences, correctly, while Z test was able to pick up 35.0% of significant differences, correctly. On raising the α level to 10%, the significant differences picked up by t-test rose to 40.4%. At α level of 15% and 20%, the percentage of correct significant differences rose to 48.9 % and 54.3% respectively. In general, Z-test was found to be picking up more significant differences than t-test.

For the Population A, B, C and D, the Results of t-test and Z-test for testing whether the set of sample means differ significantly from the respective population means, at a = 5%, is shown in Table 14. It is clear from the Pooled results shown in the table that t-test is able to pick up, correctly, the expected, non-significant differences in 96.7% of the sample means as compared to 92.5% picked up correctly by Z-test.

Table 14: Non-Significant Results of Single Mean test by different Sample size, Population and Test at a = 5%,

Population	Sample size = 9		Sample size = 13		Sample size = 20		Pooled	
	t-test	Z-test	t-test	Z-test	t-test	Z-test	t-test	Z-test
A	10	10	9	8	10	10	29	28
B	10	10	9	9	10	10	29	29
C	10	8	9	8	9	9	28	25
D	10	9	10	10	10	10	30	29
Pooled	40	37	37	35	39	39	116	111
Total	40	40	40	40	40	40	120	120
%	100	92.5	92.5	87.5	97.5	97.5	96.7	92.5

Note: For each population, 10 samples are obtained.

Table 15: Results of t-test as Compared to Z test, for the difference between two Sample means by different sample sizes and a levels (5% - 20%)

Sample	5%		10%		15%		20%	
	t-test	Z - test	t-test	Z - test	t-test	Z - test	t-test	Z - test
9A - 9A	90	88	88	86	86	82	82	76
9A - 13A	98	96	90	89	85	84	83	80
9A - 20A	99	98	97	95	93	89	88	87
13A - 13A	84	84	80	80	78	78	68	68
13A - 20A	95	94	87	85	77	76	73	72
20A - 20A	90	90	86	86	82	82	82	82
9B - 9B	90	90	90	88	88	86	86	80
9B -13B	98	97	95	94	90	88	84	84
9B - 20B	99	99	99	98	95	94	91	90
13B -13B	88	88	88	88	80	82	80	80
13B - 20B	96	96	92	92	89	89	84	84
20B - 20B	88	86	84	84	80	78	76	74
Non-Significant	1115	1106	1076	1065	1023	1008	977	957
Total test	1140	1140	1140	1140	1140	1140	1140	1140
%	97.8	97.0	94.4	93.4	89.7	88.4	85.7	83.9

The Results of t-test and Z-test, for testing the significant differences between different pairs of sample means when drawn from the same population (Population A and B), by varying a levels are shown in Table 15.

It is clear from the table that when a = 5%, t-test is able to pick up, correctly, the expected, non-significant differences in 97.8% of the sample means. This percentage rises to 85.7% when level is chosen to be 20% from 5%. When the results are seen by Z test, the percentages are not much different from those observed by t-test. At a = 20%, Z test was able to pick up 83.9% of the non-significant differences, correctly as compared to 85.7% picked up correctly by t-test.

The Results of t-test and Z-test for testing the significant differences between different pairs of sample means when drawn from the same population (Population C and D), by varying a levels are shown in Table 16.

At a = 5%, t-test was able to pick up 94.3% of the non-significant differences, correctly as against 91.8% by Z-test. At a = 20%, this percentage reduces to 80.3% and 76.3% for t-test and Z-test, respectively.

Table 16: Results of t-test as Compared to Z test, for the difference between two Sample means by different sample sizes and a levels (5% - 20%)

Sample	5%		10%		15%		20%	
	t-test	Z-test	t-test	Z-test	t-test	Z-test	t-test	Z-test
9C - 9C	76	72	68	60	60	56	56	54
9C - 13C	89	82	83	70	75	66	68	64
9C - 20C	90	84	76	73	70	70	69	67
13C - 13C	82	78	78	68	72	60	66	52
13C - 20C	87	86	81	80	76	72	70	70
20C - 20C	82	78	70	68	64	64	60	56
9D - 9D	90	90	90	90	90	88	88	84
9D -13D	100	100	98	98	98	98	96	96
9D - 20D	99	99	97	96	96	95	94	92
13D -13D	90	90	90	90	88	84	80	76
13D - 20D	100	100	98	96	92	89	88	85
20D - 20D	90	88	88	82	82	76	80	74
Non-Significant	1075	1047	1017	971	963	918	915	870
Total test	1140	1140	1140	1140	1140	1140	1140	1140
%	94.3	91.8	89.2	85.2	84.5	80.5	80.3	76.3

DISCUSSION

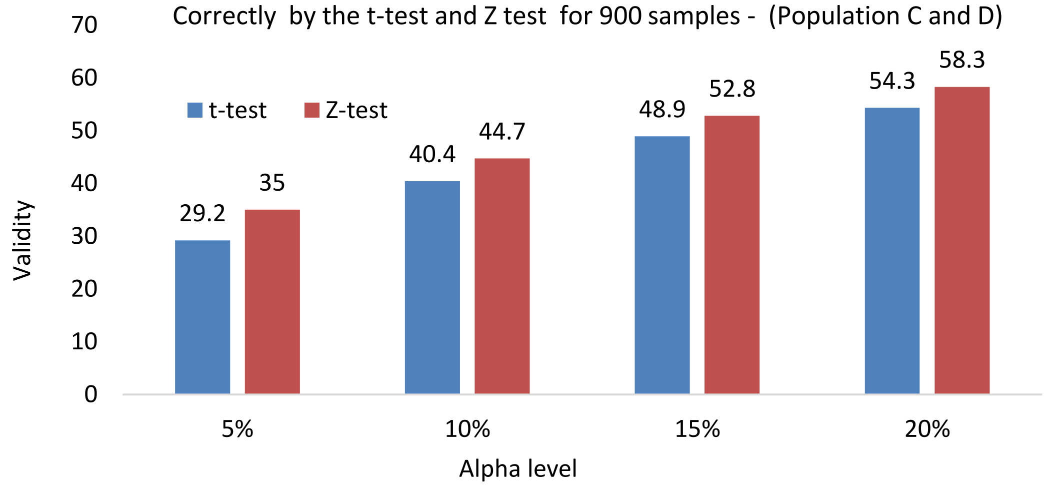
The present study was carefully designed to evaluate the performance of t-test as compared to Z-test in testing the significant or non-significant differences between two sample means. The sources of data for the study came by generating four Normal populations and then drawing 30 samples each form those populations. Thus, 120 samples, so generated allowed for 1800 mean comparison for expected significant differences and 2320 non-significant differences. Overall, 14400 comparisons were made to test for significant differences and 18240 comparisons for non-significant differences. Thus, altogether, 32640 mean tests were carried out to arrive at the validity of t-test and Z-test.

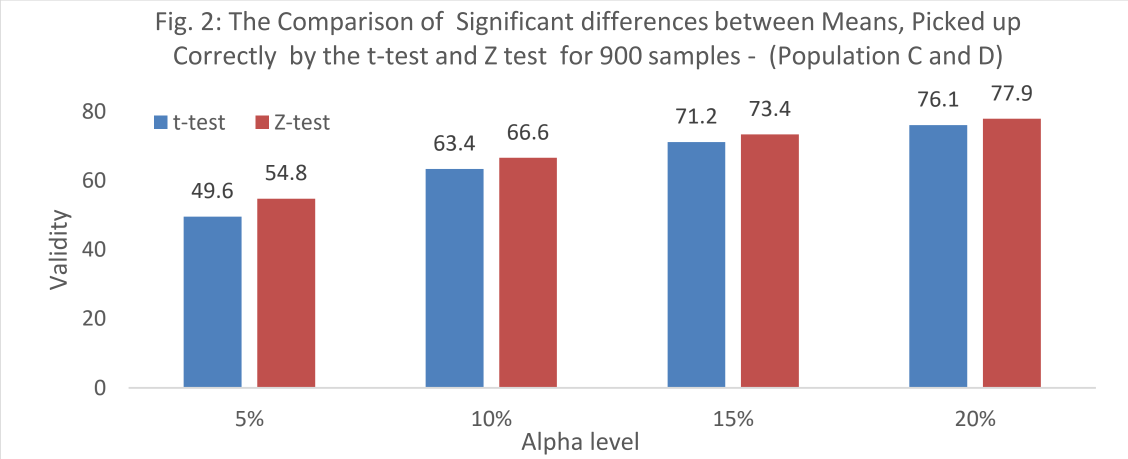
It is surprising to note that at α = 5%, t-test was able to pick up only 29.3% of the significant differences between the set of sample means of Population C and D, which is quite low suggesting that the validity of t-test in picking up the significant difference between two sample means is less than 30%. In case of Population A and B, the validity of test was observed to be relatively better and was 49.6% that is below 50%. Moreover, this conclusion has to be taken seriously as it is based on a large number namely 1800 mean comparisons.

The t-test was introduced with the understanding that for small samples, it is a better choice test as compared to Z-test for comparison between two sample means. However, the data provided in this paper does not support this view. In case of Z-test, at α = 5%, it was able to pick up higher percentage of significant differences then t-test i.e., 35% against 29.2% for the Population of C and D (Fig. 1) and 54.8% against 49.6% for the Population of A and B (Fig. 2). The superiority of Z-test is maintained even at α =20%. While considering the Population A and B, the validity was 58.3% as against 54.3%. For the Population C and D, the validity was 77.9% as against 76.1% for the Z test and t-test, respectively. This is in contradiction to the theory that for comparison of two sample means, t-test is better as compared to Z-test. Thus, the validity of the Z-test is observed to be better as compared to t-test in picking up the significant difference between the means.

In view of low validity observed in the case of α = 5%, the validity was further explored at the higher levels namely at 10%, 15% and 20%. With the rise in α levels, the validity was observed to be increasing.In case of Population C and D, at α = 20% , the validity of t-test rose to 54.3% from 29.2%, while for Population A and B, the validity rose to 76.1% from 49.6%. This suggests that there is need to revise the α level from 5% to 20% in order to get more valid mean comparisons by t-test. Based on data, it is amply clear that Z-test has relatively higher validity as compared to t-test.

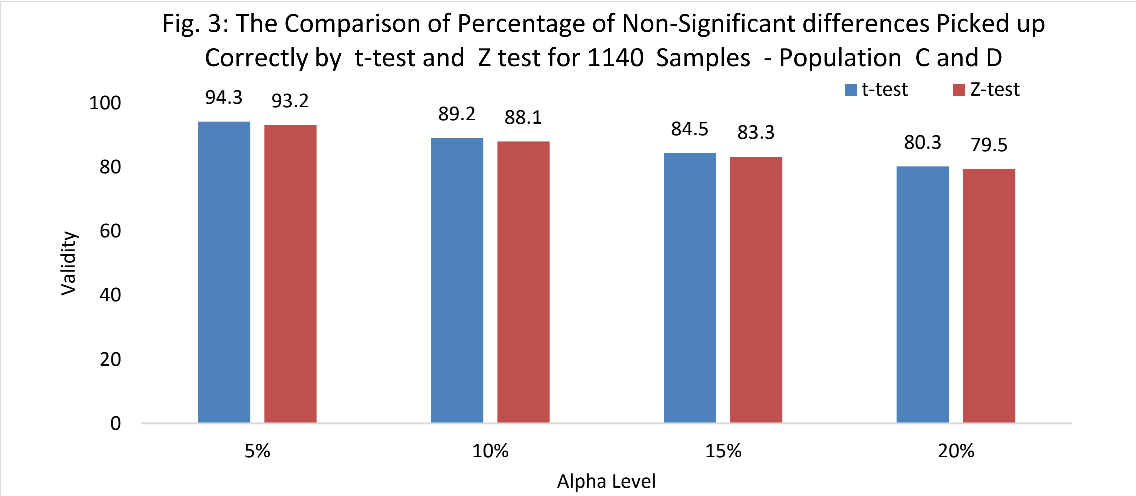
Fig.1: The Comparison of Significant differences between Means, Picked up Correctly by the t-test and Z test for 900 samples - (Population C and D)

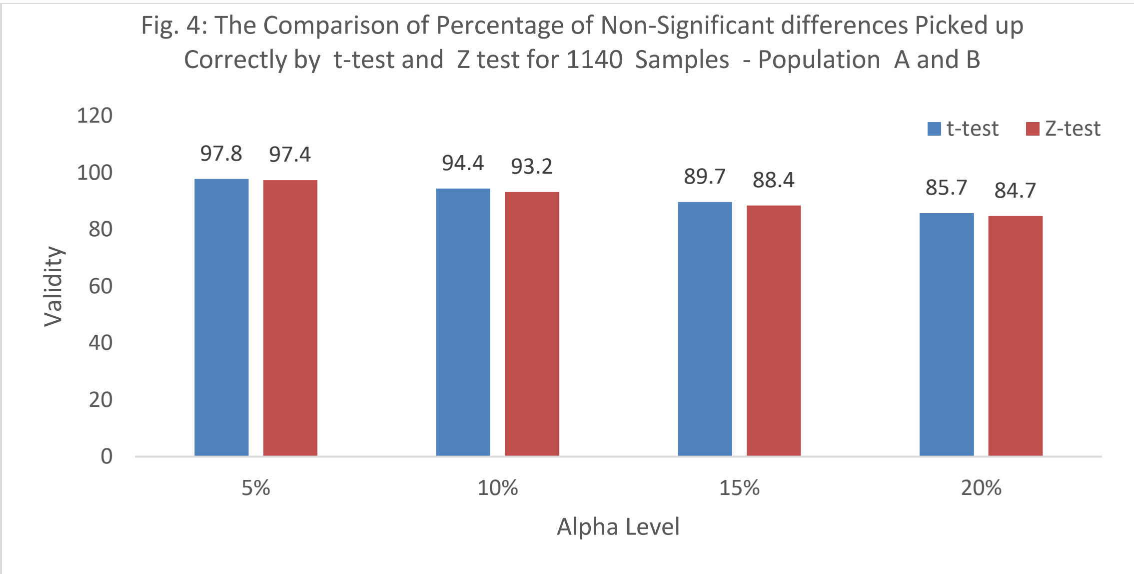




Based on findings of the present study, it can be said that t-test is not very successful in picking up the expected significant differences between two sample means. In fact, Z-test appears to be better than t-test in picking up the significant differences between two sample means.

However, it remains to be seen how does the t-test behaves when it is supposed to pick up the non-significant differences correctly between the means?





Again, based on 1140 comparisons, each, the validity of t-test was observed to more than 94% at α = 5% which is quite good. The Z-test, also performed well and the validity was more than 93%, showing not much difference from that observed for the t-test. It is further interesting to observe that the validity is maintained to be more than 80% even when α is 20% for t-test and it slightly less for Z-test (Fig.3, Fig. 4). So, it can be concluded that t-test as well as the Z test is better in picking up the non-significant differences correctly than picking up the significant differences between the means.

By continuing to use α = 5%, we may tend to reject Ho more often than required. Raising α level to 20%, while helps us in raising the percentage of picking up significant differences by t as well as Z test, it is not affecting the ability of both tests in picking up the non-significant differences, to a great extent. Thus, by raising the a level we tend to pick up more significant differences as compared to the fall in percentage in picking up the non-significant differences.

It is interesting to think why with increasing α values, t-test and Z-test are able to pick up more significant differences, correctly. The answer lies in the fact that with increasing α values, we tend to shorten the critical difference required to seek the significant difference between two means. This fact also answers to the fact that why Z-test is able to pick up more significant differences than t-test. For α = 5%, it is known that Z-test requires 1.96 as the cut off level to judge whether the difference between two means is significant or not, while for t-test, you require more than 2.0 as the critical cut off level. For example, for 20 degrees of freedom, t critical value is 2.09, implying that t-test is assigning relatively higher differences as compared to Z-test, to pick up the significant differences between two sample means. On this observation, itself, for small samples, we can justify the continuous use of Z-test in place of currently used t-test.

It is to be highlighted that even for small sample size, while Z-test is better as compared to t-test in picking up the expected significant differences correctly, it lags behind marginally with t-test in picking up the non-significant differences, correctly. Overall, there is not much gain by using t-test for small samples as advocated strongly in all text books. It is therefore suggested that we may start using Z-test even for small sample size in place of currently used t-test, to find significant or non-significant differences between two means. The major gain in doing so is the use of single level while for t-test the critical level has to be chosen based on the degrees of freedom.

CONCLUSIONS

Based on the data used in the present study, when α = 5%, it is clear that the validity of t-test remains below 50% for picking up the significant differences between two sample means. However, t-test performs far better when it comes to test the expected non-significant differences between two sample means. In this case, the validity is observed to be more than 94%,

Low validity of t-test, especially, in picking up the expected significant differences, suggests that probably, there is a need to raise the α level from 5% to 20% to improve overall validity of the t-test. This was also true in the case of Z-test.

In view of Z-test performing better as compared to t-test in picking up the significant differences, correctly, and not lagging behind much in picking up the non-significant differences between two sample means, suggests that Z-test can be used even for small sample sizes in place of hitherto used t-test.

REFERENCES

[1]. Gupta SC 2012: Fundamental of Statistics, Seventh Edition, Himalaya Publishing House; Page 19.1-19.2, 19.12-19.14.

[2]. Gupta SC, Kapoor VK 2001: Fundamentals of Mathematical Statistics, Sultan Chand & Sons; Tenth revised Edition, Page 14.1-14.3 and 14.24-14.26.

[3]. Microsoft Corporation, 2019. Microsoft Excel, Available at: https://office.microsoft.com/excel.

[4]. Snedecor GW, Cochran WG 1967; Statistical Methods, Sixth Edition; Oxford & IBH Publishing Co., Page 59-60, 100-105.

[5]. StatPlus 7.6.5.0 2021, AnalystSoft Inc. - Statistical analysis program. Version v7. See https://www.analystsoft.com/en/

Biography

Dr. Ramnath Takiar

I am a Post graduate in Statistics from Osmania University, Hyderabad. I did my Ph.D. from Jai Narain Vyas University of Jodhpur, Jodhpur, while in service, as an external candidate. I worked as a research scientist (Statistician) for Indian Council of Medical Research from 1978 to 2013 and retired from the service as a Scientist G (Director Grade Scientist). I am quite experienced in large scale data handling, data analysis and report writing. I have 57 research publications in national and International Journals related to various fields like Nutrition, Occupational Health, Fertility and Cancer epidemiology. During the tenure of my service, I attended three International conferences namely in Goiana (Brazil-2006), Sydney (Australia-2008) and Yokohoma (Japan-2010) and presented a paper in each. I also attended the Summer School related to Cancer Epidemiology (Modul I and Module II) conducted by International Agency for Research in Cancer (IARC), Lyon, France from 19th to 30th June 2007. After my retirement, I joined my son at Ulaanbaatar, Mongolia. I worked in Ulaanbaatar as a Professor and Consultant from 2013-2018 and was responsible for teaching and guiding Ph.D. students. I also taugth Mathematics to undergraduates and Econometrics to MBA students. During my service there, I also acted as the Executive Editor for the in-house Journal "International Journal of Management". I am still active in research and have published 16 research papers in past 7 years