Vol.12.Issue.2.2024 (April-June)
©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 TAKIAR Z-TEST – A BETTER OPTION THAN THE T-TEST FOR MEAN COMPARISONS AMONG SMALL SAMPLES, BELOW 30

RAMNATH TAKIAR
Flat N0.11, 3rd Floor, Building # 9, 1st Khoroo, Ulaanbaatar district, Ulaanbaatar, Mongolia- 14241
Email: ramnathtakiar@gmail.com, ramnath_takiar@yahoo.co.in
&
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.12.2.1

Ramnath Takiar

ABSTRACT

The present study is carried out to explore the relationship between the Range and the SD when drawn from the normal populations. It was shown that for normal samples, the Range to SD ratio is varying from 2.34 for the sample size of 3 to 3.70 for the sample size of 15 and 4.22 for the sample size of 30. Realizing the exponential relationship seen between the Range and the SD, a doubt is raised whether the use of uniform cut-off levels, for testing the significance differences between two sample means is justifiable or not?

The study is successful in providing a set of new cut-off levels according to selected α levels (5%, 10% and 15%) and sample size, varying from 4 to 30, utilizing the relationship of SDP with that of R95, R90 and R85, representing 95%, 90% and 85% of the range, respectively. For two sample means comparison, the study has utilized three tests namely t-test, Z-EV test and Takiar Z-test. For the Z-EV test and Takiar Z-test, the variance formula, based on the sample values as $\frac{1}{n}\Sigma(x_i - \overline{x})^2$ is used. The basic difference between Z-EV test and Takiar Z-test is that the later test makes use of cut-off values developed in the current study instead of traditional values based on the Normal table.

The study carried out 15000 mean comparisons, spread over 5 small sample sizes (4, 8, 12, 18, 24) and the 6 pairs of distinct normal populations arising from P1, P2, P3, P4, P5 and P6. In total mean comparisons, attempted, the t-test, Z-EV test, and the Takiar Z test could picked up, correctly, the expected significant differences, in 23.3%, 29.9% and 42% of the cases, respectively. The Takiar Z test, therefore, is observed to be the best in picking up correctly the expected significant differences and recommended for comparisons of sample means, for small samples, in place of t-test.

KEY WORDS: Small samples, t-test, Z-EV test, Takiar Z test, Negative Validity, Positive validity

INTRODUCTION

The t-test is widely used, especially for small samples below 30, to decide whether two sample means obtained in connection of some research study or survey are comparable or not? In the application of t-test, the sample mean is taken as the estimate of the population mean and sample variance with denominator (n-1), as the estimate of the population variance. In a few recent studies carried out (Takiar R, 2021, 2023-1, 2023-2), it was shown that for small samples, the validity of the t-test, in picking up correctly the significant differences between two sample means, is far from satisfactory. For the samples of size of 10, at α = 5%, the t-test was shown to be picking up only 31.1% of the expected significant differences between two sample means which increased to 52.0% for the sample size of 18 and 63.2% for the sample size of 24. In contrast, the Z-EV test (Z-test with estimated sample variance) picked up 39.9%, 58.2% and 68.9% correctly the significant differences between two sample means. This led to the recommendation that even for small samples, Z-EV test can be used in place of t-test.

For a normally distributed population, it is stated that 99% of the observations should lie between Mean – 3SD to Mean + 3SD. Based on this property, it can be stated that the Range to SD ratio should be around 6.0 and should not vary much with the sample size. But in a recent study (Takiar 2023-3), it was shown that the Range to SD ratio is varying exponentially according to the varying sample size (Fig. 1).

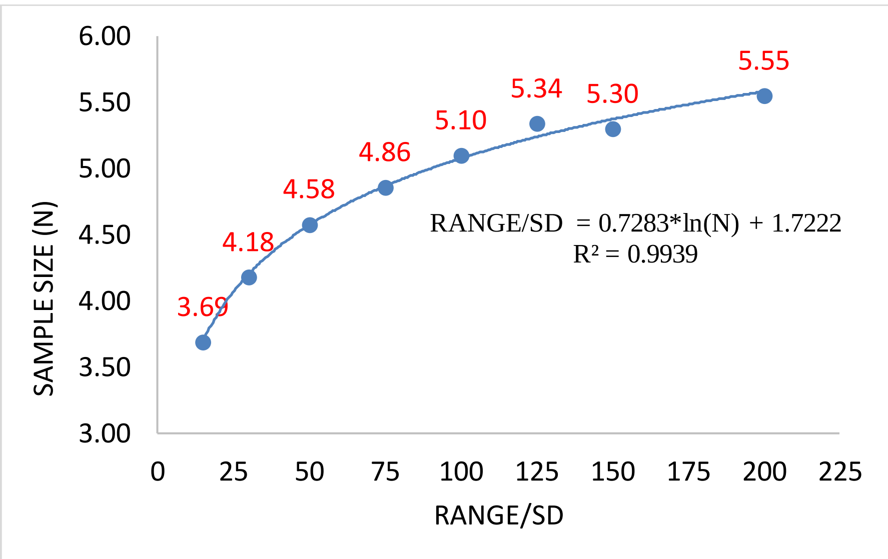

Fig. 1: The relationship between Range/SD ratio and the Sample size

According to Fig. 1, the Range to SD ratio is shown to be varying for normal samples, from 3.69 for the sample size of 20 to 4.58 for the sample size of 50 and 5.10 for the sample size of 100. Accordingly, the 99% confidence limit multiplier should be 1.85, 2.29 and 2.55 respectively, while theoretically, it is 2.58.

Based on the above finding, it is logical to think that the use of a uniform cut off level of 2.58, hitherto used for defining 99% Confidence Interval is not appropriate for normal samples and a set of fresh cut-off levels based on the sample size to be re-defined and used. In view of the recommendation made (Takiar R, 2023t) that Z-EV test can be safely used in place of t-test, even for small samples, it become necessary that for varying sample size below 30, a fresh set of Cut-off levels are defined and used. The present study is therefore designed with the following objectives.

OBJECTIVES

- To explore the relationship between the Standard Deviation and the Range for the Normal samples of selected sizes below 30.
- Utilizing the type of relationship seen between the Range and the SD, an attempt will be made to define the Cut-off levels for selected α levels and the sample sizes below 30?
- To test the validity of the Cut-off levels so developed in picking up correctly the significant differences between two sample means and compare it with that of picked up by the Z-EV test and the t-test.

MATERIALS AND METHODS

Z TEST FOR COMPARISON OF TWO MEANS

In case of two independent samples, the primary interest is to compare and decide whether two samples have comparable means or not? The statistics used thereby is:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{S} \quad \text{where } S = \sqrt{\frac{\sigma^2{}_1}{n_1} + \frac{\sigma^2{}_2}{n_2}} \quad \text{.......................... (1)}$$

For Z statistics, the sample means are taken as the estimates of the respective population means. In case of Z test, σ is assumed to be known.

Whenever, σ is unknown, particularly for small samples, below 30, its sample estimate is used in the formula (1). Such a Z test is termed as Z-EV test. For Z-EV test, the formula used for estimating the sample variance is given as

$$\text{SDP}^2 = \frac{1}{n}\Sigma(x_i - \mu)^2$$

DESCRIPTION OF THE NORMAL POPULATIONS

For the study purposes, four Normal populations are considered. The description of the populations along with their Mean and SDP are shown in Table 1.

Table 1: Description of Normal Populations with Specified Mean and SD

POPULATION	P1	P2	P3	P4
N	200	200	200	200
MEAN	55.5	44.21	65.77	76.14
SDP	16.013	11.697	17.946	12.861

SELECTION OF SAMPLES AND SAMPLE SIZE

From each of the four populations, 50 random samples with the size of 3, 6, 9, 12, 15, 18, 21, 24, 27 and 30 are generated, using the program on Excel and pooled. Thus, in total 200 samples are generated for each sample size.

DATA COLLECTED

From each sample, for the given sample size, the following statistics are collected: Sample size (n), Mean, SDP, Minimum value (MIN), Maximum value (MAX).

GENERATION OF PERCENTILE VALUES

The following percentile values are generated like P(2.5), P(97.5), P(5.0), P(95.0), P(7.5), P(92.5), from the data using the Excel function PERCENTILE.INC

TYPES OF RANGES

Based on the percentile values the following four types of Ranges are calculated.

- 100% Range = R100 = MAX – MIN
- 95% Range = R95 = P(97.5) - P(2.5)
- 90% Range = R90 = P(95.0) - P(5.0)
- 85% Range = R85 = P(92.5) – P(7.5)

DEFINTION OF RANGE TO SDP RATIOS

For each sample size**,** the following four Range to SDP ratios are calculated.

R100/SDP, R95/SDP, R90/SDP, R85/SDP

For a given sample size, the mean of 200 Range to SDP ratios is taken to represent it. It may be recalled that 50 samples each, for each sample size generated, are pooled in order to cover the variability in Range to SDP ratios arising due to difference in parameters of the different populations.

DEVELOPMENT OF REGRESSION EQUATION FOR SDP TO RANGE

The means of R95/SDP ratios, available for different sample sizes namely 3,6, 9, 12, 15, 18, 21, 24, 27 and 30, are utilized to develop a regression equation. For arriving at the regression equation, the log of the sample size is considered as X and the mean of the corresponding R95/SDP ratio is taken as Y. The regression equation, thus, developed, is utilized to define the Cut-off values for varying sample size namely from 4 to 30. Similar, exercise is attempted in case of R90/SDP and R85/SDP ratios.

DEVELOPMENT OF TABULATED VALUES FOR α (5%, 10%, 15%)

- R95/SDP values derived with the help of the regression equation for different sample sizes is taken to represent the critical values when α = 5%
- R90/SDP values derived with the help of the regression equation for different sample sizes is taken to represent the critical values when α = 10%
- R85/SDP values derived with the help of the regression equation for different sample sizes is taken to represent the critical values when α = 15%

It should be noted that the Cut-off levels generated are not based on theoretical distribution but based on model generated values.

TESTS SELECTED FOR TESTING THE SIGNIFICANCE DIFFERENCES AMONG SAMPLE MEANS

- t-test
- Z-EV test (Z with Estimated Variance from the sample values)
- Modified Z test**.** To differentiate well with that of Z-EV test, hereafter, this test will be referred as Takiar Z-test.

For the t-test, the SD with the variance formula of $\frac{1}{n-1}\Sigma(x_i - \bar{x})^2$ is used.

For the Z-EV test and Takiar Z-test, the SDP with the variance formula, based on the sample values as $\boldsymbol{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$ is used.

The basic difference between Z-EV test and Takiar Z-test is that the later test makes use of cut-off values developed in the current study instead of traditional values based on the Normal table.

VALIDITY OF T-TEST, Z-EV TEST AND TAKIAR Z-TEST WHEN MULTIPLE COMPARISONS ARE MADE

In general, the hypothesis tested for testing two sample means will be defined as follows: $H_o : m_1 = m_2$ and $H_1: m_1 \neq m_2$

WHEN SAMPLES ARE DRAWN FROM TWO DIFFERENT NORMAL POPULATIONS

In this case, it is logical to reject the Null Hypothesis. Thus, the Negative validity of the test under consideration can be defined as follows:

Negative Validity = [Number of significant differences found correctly /500] *100

Where 500 is the number of Mean Comparisons made.

WHEN SAMPLES ARE DRAWN FROM SAME NORMAL POPULATION

In this case, it is logical to accept the Null Hypothesis. Thus, the Positive validity of the test under consideration can be defined as follows:

Positive Validity = [Number of Non-significant differences found correctly /500] *100

Where 500 is the number of Mean Comparisons made.

THE SCHEME OF MEAN COMPARISONS AMONG DIFFERENT SAMPLE MEANS

The Scheme of sample Mean Comparisons when drawn from different populations is shown in the Table 2. For the Mean comparisons, two more populations namely P5 and P6 are considered. This was done with the purpose to increase the number of mean comparisons between different population samples. For the study purposes, 5 sample sizes (4, 8, 12, 18 and 24) are considered. For each sample size, 500 samples are generated. Thus, 2500 sample mean comparisons are attempted for each of the six selected pairs of populations {(P1,P2), (P1,P5), (P2,P5), (P3,P4), (P3,P6), (P4,P6)}. Overall, 15000 mean comparisons are attempted.

RESULTS

The results of comparison of populations means, for six pairs of populations, are shown in Table 3. All mean comparisons attempted are shown to be significantly different from each other.

Table 2: Scheme of Sample Mean Comparisons When Drawn from the Different Populations

	P1	P2	Mean Comparisons	Population		Mean Comparisons
Sample Size	4	4	500	P1	P2	**2500**
	8	8	500	P1	P5	**2500**
	12	12	500	P2	P5	**2500**
	18	18	500	P3	P4	**2500**
	24	24	500	P3	P6	**2500**
Total			**2500**	P4	P6	**2500**
				Total		**15000**

Table 3: Mean Comparisons Among Different Set of Populations

Population	N	Mean	SDP	Comparison	Z-VALUE	P-Value
P1	200	55.5	16.013	P1-P2	8.019	< 0.001
P2	200	44.21	11.697	P1-P5	2.5	< 0.05
P5	200	51.92	16.786	P2-P5	5.099	< 0.001
P3	200	65.77	17.946	P3-P4	7.476	< 0.001
P4	200	76.14	12.861	P3-P6	2.941	< 0.01
P6	200	70.63	17.161	P4-P6	4.078	< 0.001

Table 4: Median and Mean of Different Ratios According to Selected Sample sizes (n = 200)

Sample size	R100/SDP		R95/SDP		R90/SDP		R85/SDP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
3	2.36	2.34	2.24	2.22	2.13	2.10	2.01	1.99
6	2.93	2.91	2.74	2.73	2.56	2.55	2.39	2.37
9	3.32	3.31	3.07	3.05	2.81	2.79	2.62	2.76
12	3.52	3.57	3.24	3.24	2.91	2.91	2.60	2.58
15	3.64	3.70	3.35	3.33	2.93	2.95	2.64	2.60
18	3.80	3.82	3.40	3.42	3.00	3.01	2.72	2.69
21	3.87	3.90	3.48	3.46	3.03	3.02	2.75	2.73
24	3.99	4.03	3.56	3.52	3.07	3.05	2.75	2.73
27	4.14	4.14	3.57	3.57	3.09	3.09	2.72	2.72
30	4.21	4.22	3.63	3.63	3.17	3.16	2.80	2.78
r	-	**0.998**	-	**0.989**	-	**0.973**	-	**0.883**
b	-	**1.857**	-	**1.374**	-	**0.975**	-	**0.679**
a	-	**1.49**	-	**1.66**	-	**1.758**	-	**1.825**

The Median and Mean values of Range to SDP ratios by varying Sample size is shown in Table 4. The ratios shown in the table are: R100/SDP, R95/SDP, R90/SDP, and R85/SDP. In addition, the Correlation coefficient (r), Slope (b) and Intercept (a) are shown in the table. The Log values of Sample size are taken as X values and the corresponding average values of ratios are taken as Y values.

For R100/SDP, for the sample size of 3, the mean ratio is observed to be 2.34 and it increased to 4.22 for the sample size of 30. The corresponding mean ratios for R95/SDP are observed to be 2.22 and 3.63, respectively. For R90/SDP, it changed to 2.1 to 3.16 and for R85/SDP, it is observed to be 1.99 and 2.78, respectively. For all the four Range ratios, the mean ratios are observed to be increasing with increasing sample size. The closeness in mean and median values for all the ratios suggests the lack of skewness in the distribution of selected four Range to SDP ratios.

Table 5: Cut-Off Level According to Selected α Level

N	α Level		
	5%	10%	15%
4	1.244	1.173	1.117
5	1.310	1.220	1.150
6	1.365	1.258	1.177
7	1.411	1.291	1.199
8	1.450	1.319	1.219
9	1.486	1.344	1.236
10	1.517	1.367	1.252
11	1.545	1.387	1.266
12	1.571	1.405	1.279
13	1.595	1.422	1.291
14	1.617	1.438	1.302
15	1.638	1.452	1.312
16	1.657	1.466	1.321
17	1.675	1.479	1.330
18	1.692	1.491	1.339
19	1.709	1.502	1.347
20	1.724	1.513	1.354
21	1.738	1.524	1.361
22	1.752	1.533	1.368
23	1.766	1.543	1.375
24	1.778	1.552	1.381
25	1.790	1.560	1.387
26	1.802	1.569	1.393
27	1.813	1.577	1.398
28	1.824	1.584	1.404
29	1.835	1.592	1.409
30	1.845	1.599	1.414

The correlations ranged from 0.998 for R100/SDP to 0.883 for the ratio of R85/SDP. The Slope values for the selected four Range ratios are observed to be 1.857, 1.374, 0.975 and 0.679, respectively. The intercept values are also shown in the table. A high correlation of 0.883 to 0.998, suggests that the model fitted are good and can be used for generating different cut-off levels for the sample size of 4 to 30.

The Regression equations obtained, for the selected three Range ratios, are utilized to generate different cut-off levels according to varying sample size and are shown in Table 5.

The table provides the Cut-off level for three α levels namely 5%, 10% and 15%. In generations of above Cut-off levels, it is assumed that the distribution is symmetric and the range is distributed equally among both the sides of the mean. As expected, the Cut-off level decreased with the increase in α levels. For comparing two sample means with different n say n_1 and n_2, the Cut-off level should be seen for $\frac{(n_1 + n_2)}{2}$. In case, a fraction is obtained, it should be rounded off to the nearest integer and that integer should be taken to view the Cut-off level.

VALIDITY OF TAKIAR Z TEST IN RELATION TO T-TEST AND Z -TEST

The 500 samples of varying sample sizes (4, 8, 12, 18 and 24) are generated from each of the Population namely P1, P2, P3, P4, P5 and P6. The sample means are compared for six pairs of populations namely (P1,P2), (P1,P5), (P2,P5), (P3,P4), (P3,P6) and (P4,P6). The results obtained from the significant tests are summarized in Fig. 2 to Fig. 7

For the sample size of 4 (Fig. 2), at α = 5%, the t-test could pick up correctly only the 10% of the expected significant differences as against 24.2% picked up by Z-EV test and 44.0% by the Takiar-Z test. At α = 10% and 15%, the performance of the t-test noted to be 17.9% and 25.3%, respectively. which can be considered as a reflection of low validity of the t-test. In comparison, Z-EV test performed better. In all the three tests, Takiar Z test performance was the best and the test could pick up correctly 44.0% to 48.7% of the expected significant differences correctly.

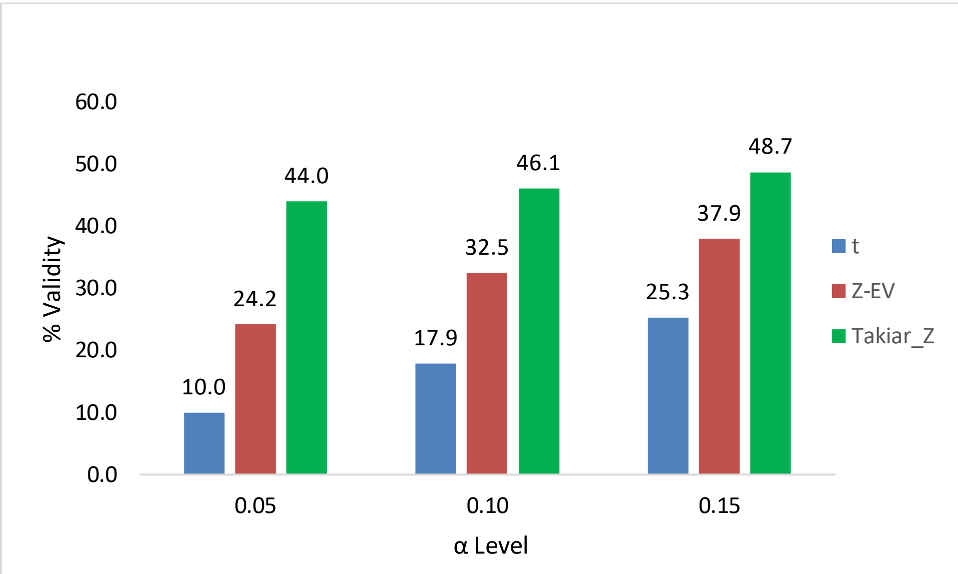

Fig. 2: The Negative Validity By the Significance tests and the α Levels - Sample size of 4 - Pooled for Population P1- P6 (n=3000)

For the sample size of 8 (Fig. 3), at α = 5%, the t-test could pick up correctly only the 16.2% of the expected significant differences, which rose to 25.7% and 33.0% for α = 10% and α = 15%, respectively. The corresponding figures for Z-EV test are observed to be 23.7%, 32.5% and 39.2%. For the Takiar Z test, the corresponding figures are observed to be 38.6%, 42.9% and 47.4%, much ahead of the t-test and the Z-EV test.

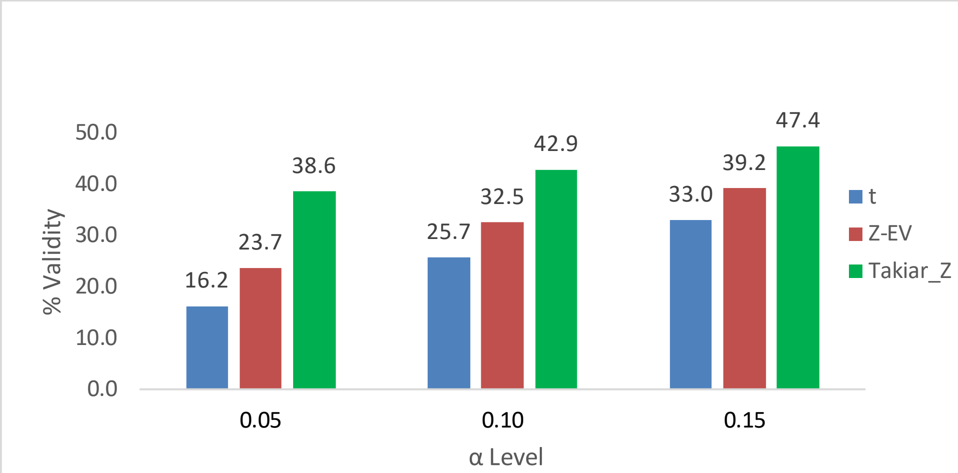

Fig. 3: The Negative Validity By the Significance tests and the α Levels - Sample size of 8 - Pooled for Population P1- P6 (n=3000)

For the sample size of 12 (Fig. 4), at α = 5%, the t-test could pick up correctly only the 21.7% of the expected significant differences, which rose to 31.9% and 39.2% for α = 10% and α = 15%,respectively. The corresponding figures for Z-EV test are observed to be 26.6%, 36.8% and 43.1% while for the Takiar Z test, the corresponding figures are observed to be 38.9%, 44.5% and 48.9%, again, much ahead of the t-test and the Z-EV test.

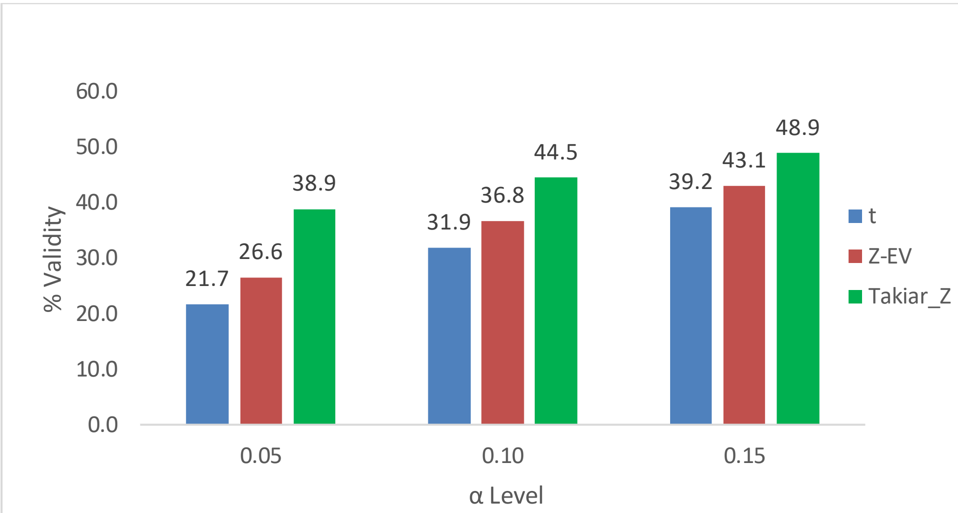

Fig. 4: The Negative Validity By the Significance tests and the α Levels - Sample size of 12 - Pooled for Population P1- P6 (n=3000)

For the sample size of 18 (Fig. 5), at α = 5%, the t-test could pick up correctly only the 30.4% of the expected significant differences, which rose to 41.2% and 48.5% for α = 10% and α = 15%, respectively. The corresponding figures for Z-EV test are observed to be 34.4%, 44.0% and 50.9%. For the Takiar Z test, again the performance was better and the corresponding figures are observed to be 42.3%, 49.3% and 55.4%,

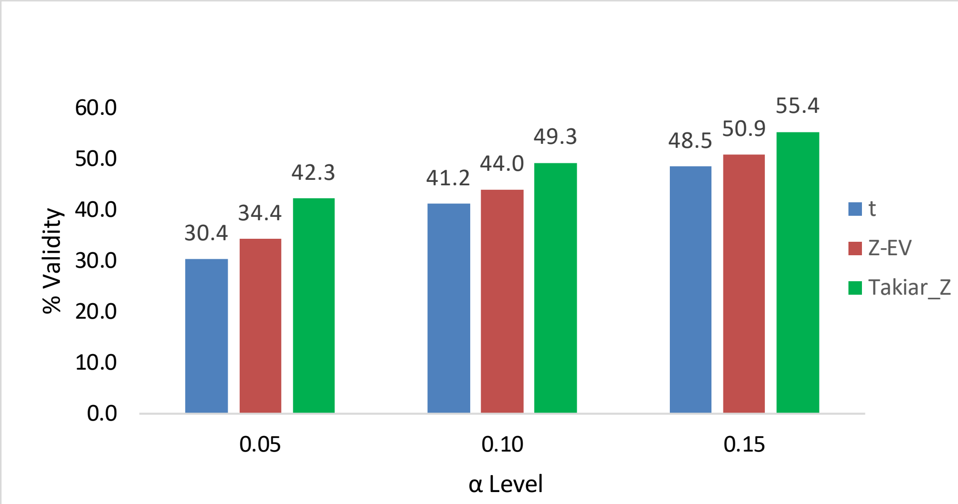

Fig. 5: The Negative Validity By the Significance tests and the α Levels - Sample size of 18 - Pooled for Population P1- P6 (n=3000)

For the sample size of 24 (Fig. 6), for α = 5%, 10% and 15%, the percentage of the expected significant differences picked up correctly by the t-test are observed to be 37.8%, 48.8% and 55.8%, respectively.

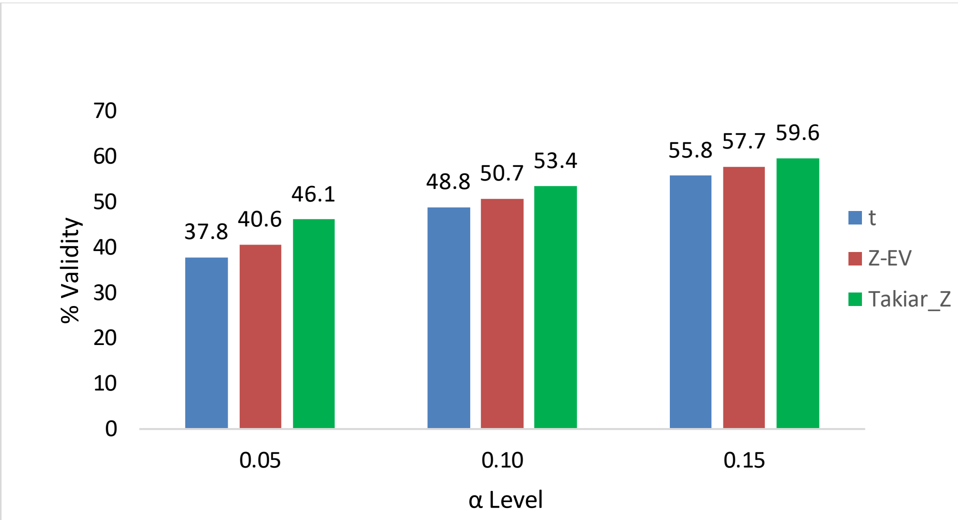

Fig. 6: The Negative Validity By the Significance tests and the α Levels - Sample size of 24 - Pooled for Population P1- P6 (n=3000)

For Z-EV test, the corresponding figures are observed to be 40.6%, 50.7% and 57.7%, respectively. For Takiar Z test, the corresponding figures are observed to be 46.1%, 53.4% and 59.6%. Again, the performance of the Takiar Z test is observed to be better than the other two tests.

The results obtained for the sample size of 4, 8, 12, 18 and 24, discussed above, are pooled and shown in Fig. 7. At α = 5%, 10% and 15%, the percentage of the expected significant differences picked up correctly by the t-test are observed to be 23.3%, 33.1% and 40.4%, respectively. For Z-EV test, the corresponding figures are observed to be 29.9%, 39.3% and 45.7%, respectively. For Takiar Z test , the corresponding figures are observed to be 42.0%, 47.2% and 52.0%, respectively. It is clear from the data that Takiar Z test is performing better as compared to other two tests.

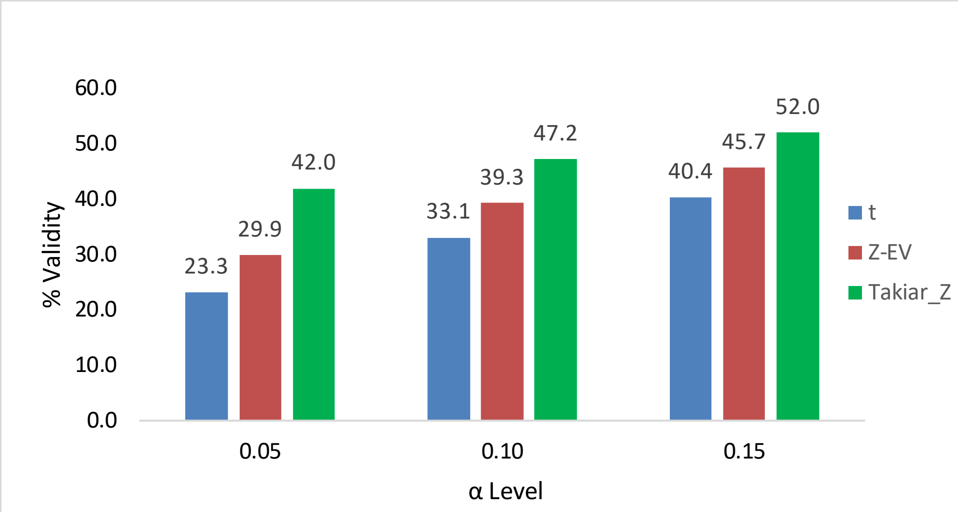

Fig. 7: The Negative Validity By the Significance tests and the α Levels - All Samples - Pooled for Population P1- P6 (n=15000)

The Positive validity obtained individually for the sample size of 4, 8, 12, 18 and 24, are not shown. The pooled positive validity for them is shown in Fig. 8. At α = 5%, 10% and 15%, the percentage of the expected non-significant differences picked up correctly by the t-test are observed to be 95.5%, 90.5% and 85.5%, respectively. For Z-EV test, the corresponding figures are observed to be 91.7%, 86.3% and 80.7%, respectively. For Takiar Z test , the corresponding figures are observed to be 83.1%, 79.1% and 75.2%, respectively. In case of Positive validity, t-test performs better as compared to other two tests which is not surprising as the t-test has tendency to accept Ho in at least 80% of the cases even when the samples are drawn from two different populations and compared.

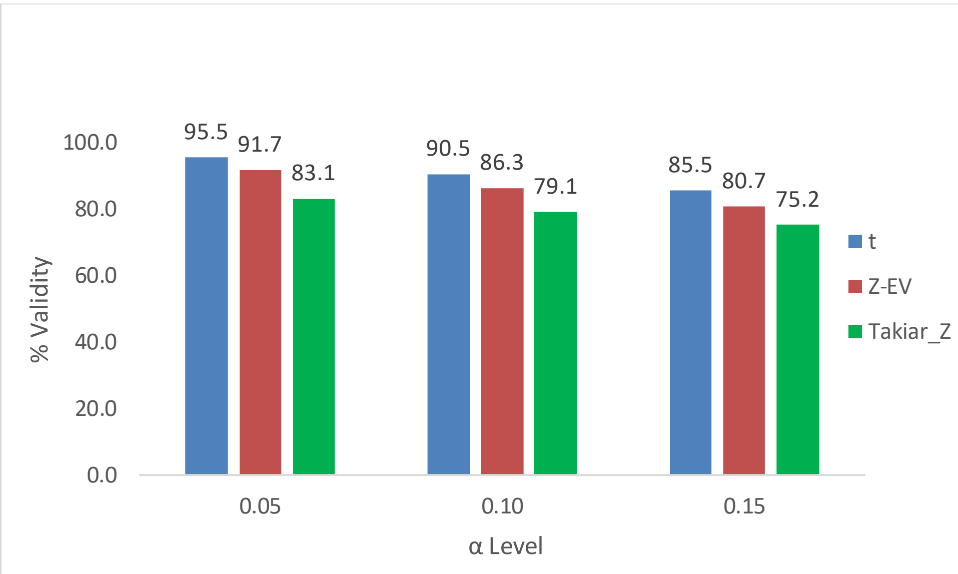

Fig. 8: Positive Validity By the Significance tests and the α Levels - All Samples - Pooled for Population P1- P6 (n=15,000)

The pooled results of negative and positive validity for the tests are shown in Fig. 9. This will give an idea about the overall capacity of the tests to pick up correctly the mean differences.

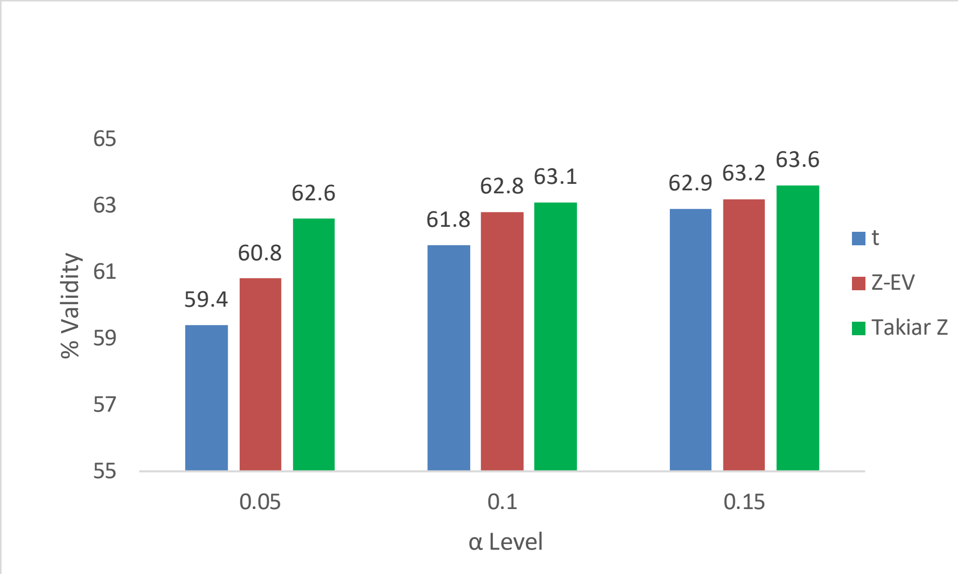

Fig. 9: Validity By the Significance tests and the α Levels - All Samples - Pooled for Population P1-P6 (n=30,000)

Based on the figures provided in Fig. 9, it can be concluded that Takiar Z test is the best test. Thus, in pair of samples where significance and non-significance are equally probable, Takiar Z test stands as the best test followed by the Z-EV test.

DISCUSSION

The study is successful in providing a set of new cut-off levels for sample size varying from 4 to 30 utilizing the relationship of SDP with that of R95, R90 and R85, representing 95%, 90% and 85% of the range, respectively.

To test the validity of the Cut-off points suggested for varying sample size, 2500 mean comparisons are attempted between each pair of the population samples, spread over 5 sample sizes namely 4, 8, 12, 18, 24, arising from {(P1,P2), (P1,P5), (P2,P5), (P3,P4), (P3,P6), (P4,P6)}. Thus, the observations discussed are based on 15000 mean comparisons.

The Takiar Z test, showed, uniformly, a higher validity in picking up the significant differences between two sample means when drawn from different normal populations. For α = 5%, the validity of the Takiar Z test is observed to be 42%, much ahead than seen in the case of t-test (21.3%) and Z-EV test (29.9%). For α = 10%, the validity is seen to be 33.1%, 39.9% and 47.2% for the t-test, Z-EV test and Takiar Z test, respectively. So, based on the validity, it can be said that Takiar Z test is the best test for small samples, for testing the significant differences between two sample means.

In case of Positive Validity, at α = 5%, when pooled for all sample sizes and samples drawn from all pairs of populations, the t-test showed 95.5% as compared to 91.7% and 83.1% for Z-EV test and Takiar Z test, respectively. The fact that at α = 5%, the test can pick up only 21.3% of expected significant differences, points out that the t-test has the tendency to accept H_0, in around 80% of the cases even when the samples are drawn from two known different populations. This also points out that irrespective of whether samples are drawn from same population or different populations, in about 80% of the cases, the t-test accepts H_0. So, there is no point in testing the capacity of the t-test, in picking up non-significant differences, correctly, in comparison to other two tests, when the samples are drawn from the same population.

In comparison to t-test and Z-EV test, the Takiar Z-test is adjudged as the best when results are pooled for negative and positive validity and then comparisons are made among the three tests. This shows that the gain in negative validity for the Takiar Z test is more than the loss in the positive validity as compared to other two tests. Thus, in a number of mean comparisons, if you give equal probability to getting significant or non-significant results, the Takiar Z tests stands as the best. Hence, the Takiar Z test can be advocated for small samples in place of t-test.

It is interesting to see what happened to positive validity of three tests when adjusted for false positive validity rate, reflected in Fig. 7. The False positive validity rate can be given as

False Positive Validity Rate = 100 – True Negative Validity Rate and

Adjusted Positive Validity = Positive Validity – False Positive Validity

The Positive validity, adjusted for False Positive Validity Rate, by the selected tests and α levels, is shown in Fig. 10. Again, the adjusted Positive Validity appears to be higher for the Takiar Z-test as compared to other two tests. It is difficult to explain why a test which is low in negative validity is the best for positive validity. An intuitive explanation is that probably the Cut-off levels assigned for the test are higher than the expected to discriminate between significance and non-significance results.

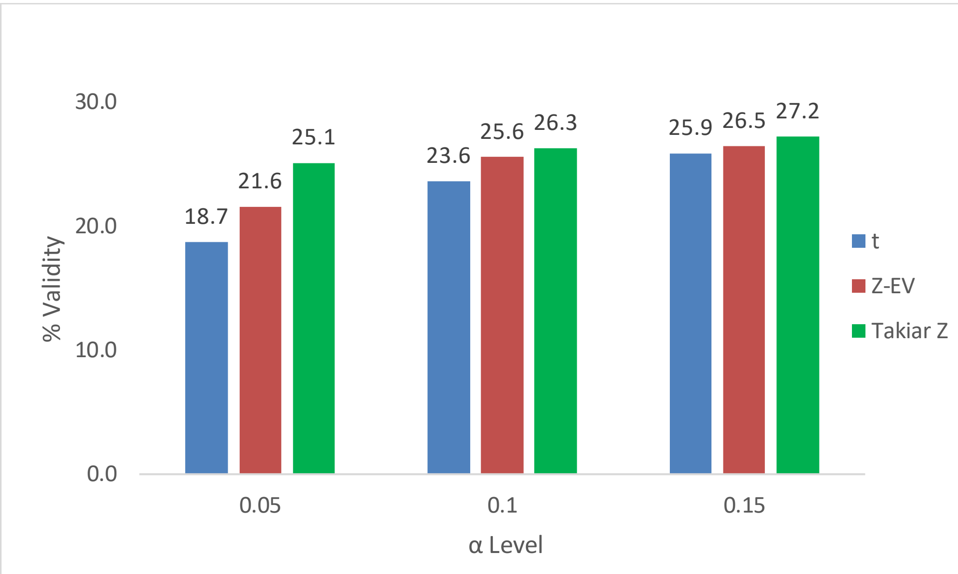

Fig. 10: Adjusted Positive Validity By the Significance tests and the α Levels - All Samples - Pooled for Population P1- P6 (n=15,000)

One of the reasons, for t-test to accept H_0, more frequently is the way it defines the standard deviation that is with (n-1) as the denominator. It is shown that the Variance with (n-1) as the denominator, against the popular claim, does not give an unbaised estimate (Takiar R, 2022). In fact, this way of defining the variance tend to consistently overestimate the population variance by 17% leading to lowering of the calculated t-values and thereby leading to acceptance of H_0, in large numbers than required. The t-test was introduced with the claim that it is a better test than Z-test especially for small samples but the study results do not support this view. Further, for small samples, the Z-EV test has been shown to be performing better than the t-test.

The study has brought clearly that even for small samples, Takiar Z-test is the best when compared to the t-test and Z-EV test for testing the significant differences between sample means.

CONCLUSIONS

- The results listed below are based on 15,000 mean comparisons, spread over 5 small sample sizes (4, 8, 12, 18, 24) when drawn from the 6 pairs of distinct populations {(P1,P2), (P1,P5), (P2,P5), (P3,P4), (P3,P6), (P4,P6)}.
- An equal number (15,000) of mean comparisons are also made when distinct samples are drawn and compared from similar populations like {(P1,P1), (P2,P2), (P3,P3), (P4,P4), (P5,P5), (P6,P6)}.
- For α = 5%, the t-test picked up only 23.3% of the expected significant differences and 95.5% of the expected non-significant differences, correctly. When pooled together, the t-test delivered the correct results in 59.4% of the cases.
- For α = 5%, the Z-EV test picked up only 29.9% of the expected significant differences and 91.7% of the expected non-significant differences, correctly. When pooled together, the Z-EV test delivered the correct results in 60.8% of the cases.
- The Cut-off levels are suggested in Table 5, for sample size varying from 4 to 30 and for three α levels namely 5%, 10% and 15%, to be utilized by the Takiar Z test for arriving significant or non-significant differences between two sample means.
- For α = 5%, the Takiar Z test picked up only 42.0% of the expected significant differences and 83.1% of the expected non-significant differences, correctly. When pooled together, the Takiar Z test delivered the correct results in 62.6% of the cases.
- Based on 30,000 mean comparisons, the Takiar Z test is shown to be the best as compared to t-test and Z-EV test, in delivering the correct comparisons results.

RECOMMENDATIONS

- In case of small samples, for mean comparisons, the use of Takiar Z-test is advocated, in place of t-test, with α=10%
- For Cut-off levels, according to the selected sample size, the values provided in Table 5 should be referred.

REFERENCES

[1]. Takiar R (2021): The Validity of $\boldsymbol{t}$-test and $\boldsymbol{Z}$-test for Small One sample and Small Two Sample test, *Bulletin of Mathematics and Statistics Research*, Vol. 9(4), page 42-57.

[2]. Takiar R (2022): Sample Variance – Is it Really Unbiased Estimate of the Population Variance? , *Bulletin of Mathematics and Statistics Research,* Vol. 10(1), page 21-30.

[3]. Takiar R (2023-1): The Validity of t-test, Mann-Whitney test and Z test for Testing Significant differences between two Sample Means When Sample size is 10 or below, *Bulletin of Mathematics and Statistics Research*, Vol. 11(2), page 1-15.

[4]. Takiar R (2023-2): The Validity of t-test, Mann-Whitney test and Z Test for Testing Significant Differences Between Two Sample Means When the Sample Size is Between 10 and 30, *Bulletin of Mathematics and Statistics Research*, Vol. 11(3), page 1-12.

[5]. Takiar R (2023-3): The Relationship between the SD and the Range and a method for the Identification of the Outliers, *Bulletin of Mathematics and Statistics Research*, Vol. 11(4), page 62-75.

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 Scientist G (Director Grade Scientist). I am quite experienced in large scale data handling, data analysis and report writing. I have 72 research publications , with 1250 citations to my credit, published 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 the 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 11 research papers during 2021-24.