



<http://www.bomsr.com>

Email: editorbomsr@gmail.com

RESEARCH ARTICLE

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
2348-0580

THE VALIDITY OF MANN-WHITNEY TEST FOR COMPARING TWO SAMPLE DISTRIBUTIONS FOLLOWING UNIFORM OR EXPONENTIAL DISTRIBUTION

RAMNATH TAKIAR

Apartment no. 11, Building 9, 3 rd Floor, Mangol Expopt Town, 1st Khorro, Olympic Street
Sukhbaatar District, Ulaanbaatar, Mongolia.

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.11.3.45](https://doi.org/10.33329/bomsr.11.3.45)



Ramnath Takiar

ABSTRACT

In case of small samples (< 30) when interest is to compare two sample distributions and when the non-normality is suspected in the samples drawn, a non-parametric test is recommended. In one of my recent studies, in case of small Normal samples, below 10, it was shown that at 5% and 10% α levels, the Mann Whitney test is not suitable for comparisons of distributions and has lower validity, below 45%. However, it remains to be seen that how does Mann Whitney test perform, in terms of validity, when the samples are drawn from known non-normal distributions and compared? For the study purposes, the four sets of Uniform and Exponential populations of size 200 are generated. From each population, 500 Random samples of size 15, 20 and 25 are randomly selected. The scheme of comparison selected allowed 1000 comparisons between any two sample distributions. For the study purposes, the validity is defined as the ability of the Mann Whitney test, in terms of percentage, to pick up correctly the expected significant differences between the sample distributions. The analysis of the data indicated that 1) At $\alpha = 5\%$, the Mann Whitney showed a very low validity, below 15% for the selected sample size of 15, 20 and 25. 2) When α level is increased to 15%, even then the validity remained below

30%. 3) The test is not at all suitable when dealing with the samples of Uniform and Exponential distribution. 4) In general, the statement that the non-parametric tests are suitable for any non-normal sample comparisons is not valid and should be used with due caution. 5) The suitability of the test for the samples of other non-normal populations, namely, populations following other than Uniform and Exponential distribution, need to be explored.

Keywords: Non-parametric methods, Uniform population, Exponential population, Mann Whitney test, Validity.

INTRODUCTION

In case of large samples (more than 30), to test the significance difference between two sample means, Z test or Normal test is available to us. However, for small samples (< 30) when interest is to test the significance difference between two sample means, t-test is advocated. For application of t test, it is assumed that the samples drawn are coming from Normal populations. However, when the non-normality is suspected in the samples drawn, a non-parametric test is recommended. Compared to a parametric test like t-test or Z-test, a non-parametric test, does not make any assumption regarding the distribution of the populations. Therefore, a non-parametric test, is also called sometimes as a distribution free method (Gupta SC and Kapoor VK, 2001, Gupta SC, 2012). By non-normality, it is expected that the distributions of samples are either skewed or have large variances. In one of my recent studies (Takiar 2023), in case of small Normal samples, 10 or below, it was shown that at 5% and 10% α levels, the Mann Whitney test is not suitable for comparisons of distributions and has lower validity, below 45%. However, it remained to be seen that how does Mann Whitney test performs, in terms of validity, when the samples are drawn from known non-normal distributions and compared?

OBJECTIVES

- To evaluate the performance of Mann Whitney test when the samples are drawn from non-normal populations and intended to be compared?
- How does the validity of the Mann Whitney test vary with the varying sample size?

MATERIAL AND METHODS

SELECTION OF NON-NORMAL POPULATIONS

For the study purposes, it was decided to consider two non-normal populations, namely, the populations following the Uniform and the Exponential distribution.

UNIFORM DISTRIBUTION

It is a probability distribution in which all possible outcomes are equally likely. The probability distribution can be discrete or continuous. In case of discrete distribution, the outcomes are discrete while in the case of continuous distribution, the outcomes are continuous and infinite. For the study purposes, the continuous Uniform distribution is considered. The probability density function of the Uniform distribution is given by

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

= 0 elsewhere

Thus, the distribution is known to have two parameters namely a and b where “a” is the minimum value and “b” is the maximum value. The mean and median of the distribution is given by the formula Mean = Median = $\frac{(a + b)}{2}$ and Variance = $\frac{(b - a)^2}{12}$.

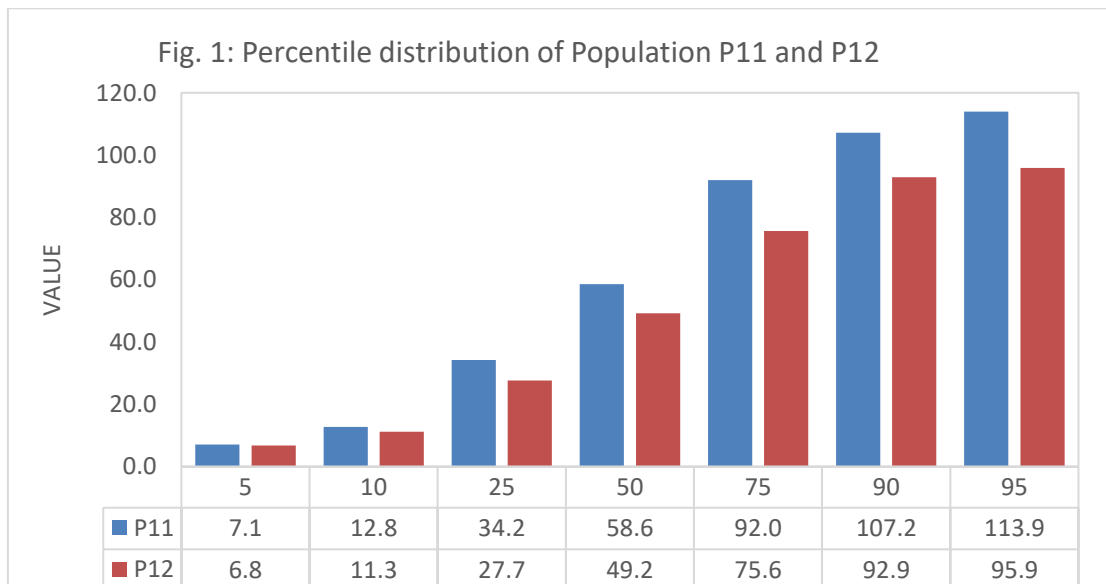
EXPONENTIAL DISTRIBUTION

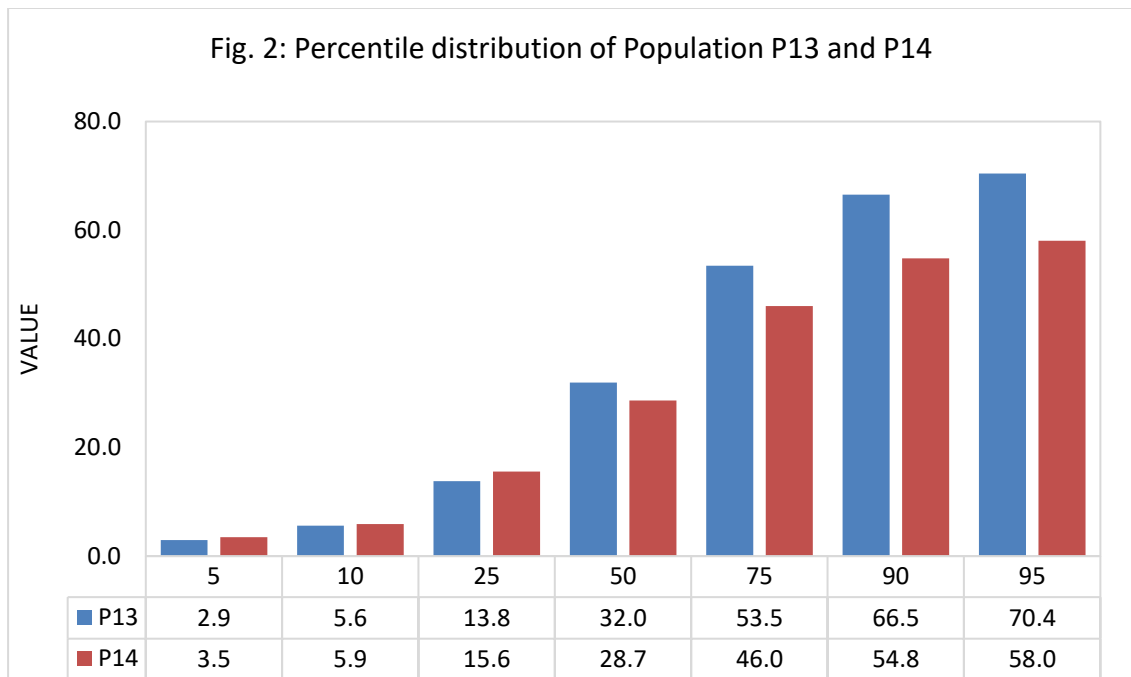
It is a continuous probability distribution in which occurring of a specific event is seen in relation to time. In a drug study, the amount of drug absorbed in relation to varying time is assessed when a specified amount of drug was injected in the experimental subjects. In an industry, in a production unit, it is used to find out the time taken to receive a defective material. In a hospital study, it is the number of Covid patients arriving in a specified unit of time.

The probability density function of an Exponential distribution is given by $y = ab^x$ where, $a > 0$ and as x increases linearly, y increases Exponentially. Whenever $b > 1$, the function increases exponentially and represented as a growth function. In case of $b < 1$, the function decreases exponentially and represented as a decay function. It is to be noted that when $x = 0, y = a$ which is known as the initial value or the base value of the function.

GENERATION OF UNIFORM POPULATIONS

The four sets of Uniform populations of size 200 are generated with the help of the function key “Random number Generation” provided in StatPlus 7.6.5. The input required in sequence, in this connection are: Number of New Variables (1), Random Number Count (200), Lower Bound, Upper Bound. For generation of four Uniform populations, the following four ranges were used as inputs: 0-120, 0-100, 0-75 and 0-60. The populations so generated are termed as P11, P12, P13 and P14. The percentile distributions of P11 and P12 are shown in Fig.1 while the percentile distributions of P13 and P14 are shown in Fig. 2.



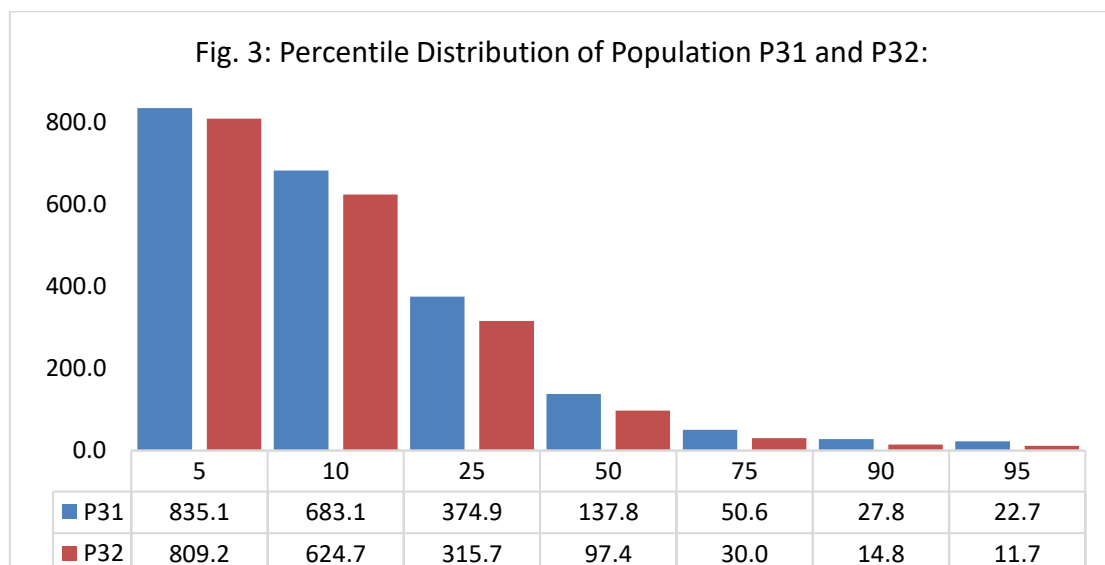


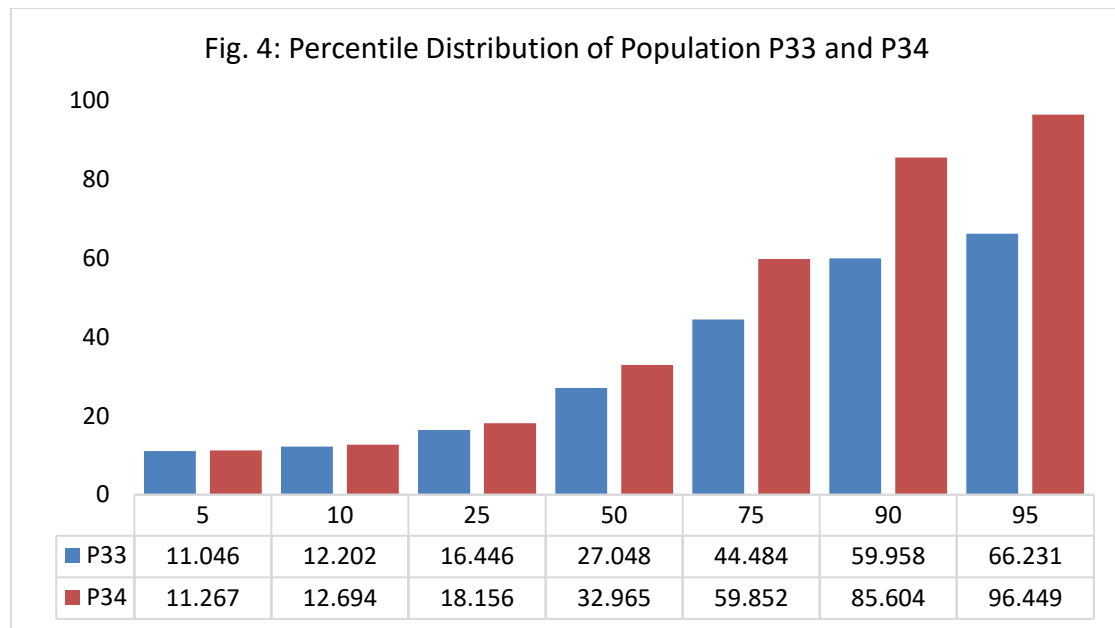
GENERATION OF EXPONENTIAL POPULATIONS

The four sets of Exponential populations of size 200 (n) are generated on Excel by selecting different sets of “a” and “b” values as shown in Table 1. The populations so generated are termed as P31, P32, P33 and P34. Further, it is intended to compare the samples of P31 and P32. Besides, the samples of P33, P34 are also compared. The percentile distributions of P31 and P32 are shown in Fig.3 while the percentile distributions of P33 and P34 are shown in Fig. 4.

Table 1: Description of Parameters used in Generation of Exponential Populations

Parameter	P31	P32	P33	P34
a	1000	1000	10	10
b	-0.002	-0.00235	0.01	0.012
n	200	200	200	200





SAMPLE SELECTION

The Scheme of sample selection by the population and the size is shown in Table 2. From each population, 500 Random samples of size 15, 20 and 25 are generated using the key “Random Sample” available with StatPlus 7.6.5. Thus, for each Population, 1500 samples are drawn, as shown below in Table 2.

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

Population	Sample size			
	15	20	25	TOTAL
P11	500	500	500	1500
P12	500	500	500	1500
P13	500	500	500	1500
P14	500	500	500	1500
P31	500	500	500	1500
P32	500	500	500	1500
P33	500	500	500	1500
P34	500	500	500	1500
Total	4000	4000	4000	12000

COMPARISONS OF POPULATION DISTRIBUTIONS

In order to assess whether the distribution of P11 is significantly different from P12, the frequency distributions are obtained against the pooled percentile values of P11 and P12 and compared by the χ^2 test. Similar approach is used to compare the significant differences between the population distributions of P13 and P14. The distributions of P31 and P32 besides the distributions of P33 and P34 are also compared using the χ^2 test.

SCHEME OF SAMPLE COMPARISONS

As per the plan, 500 samples of given size, are drawn from each population. However, for a given sample size, the distributions are compared between the samples of P11 and P12 on one hand and P13 and P14 on the other hand. By changing the sequence of samples in P12, 500 more comparisons are made between P11 and P12. Thus, in total, 1000 distribution comparisons are attempted between P11 and P12. Proceeding in a similar way, again a total of 1000 distribution comparisons are attempted between P13 and P14. Similar approach is used for comparisons of sample distributions of P31 and P32 on one hand and P33 and P34 on the other hand. The scheme of comparisons by the type of the population and the sample size is shown in Table 3.

Table 3: Scheme of Comparisons by Sample size and Populations

Distribution	Sample size	Populations	Comparisons	Populations	Comparisons
Uniform Distribution	15	P11P12	500	P13P14	500
		P11P12A	500	P13P14A	500
	20	P11P12	500	P13P14	500
		P11P12A	500	P13P14A	500
	25	P11P12	500	P13P14	500
		P11P12A	500	P13P14A	500
Exponential Distribution	15	P31P32	500	P33P34	500
		P31P32A	500	P33P34A	500
	20	P31P32	500	P33P34	500
		P31P32A	500	P33P34A	500
	25	P31P32	500	P33P34	500
		P31P32A	500	P33P34A	500

P12A – Changed sequence of samples of P12; P14A-Changed sequence of P14

P32A – Changed sequence of samples of P32; P34A-Changed sequence of P34

SIGNIFICANCE AND α LEVELS

In my earlier study (Takiar, 2023), at 5% α level, the validity of Mann Whitney test was shown to be very low. Further, the validity was seen to be rising with the rising α levels. Accordingly, for the present study, three α levels are chosen namely 5%, 10% and 15%. For testing the significant

differences between the distributions of two samples, when drawn from non-normal distributions, Mann Whitney test is used.

VALIDITY OF MANN WHITNEY TEST

For the present study, according to selected scheme of comparison, 1000 comparisons can be made between each two population samples of size 15, 20 and 25. In this case, the formulated Null Hypothesis is "The sample distributions are comparable" must be rejected and the alternative Hypothesis that the sample distributions are significantly different from each other must be accepted. A higher percentage of significant differences between the sample distributions, based on the predefined cut-off level, say above 80%, suggests the existence of the higher validity of the Mann Whitney test while a lower percentage than 80% will suggests that the validity of the test is lower.

ANALYSIS OF THE DATA

For simultaneous comparisons of the distributions of all the 500 samples, by the Mann Whitney test, SPSS program, 2023 version, is utilized. The results obtained by the test, at the given three α levels, are also compared to see whether with the rise in α level, there is a rise in the validity of the test.

RESULTS

Test of significance carried out indicated that in case of Uniform distribution, the distribution of P11 differed significantly with that of P12, so also the distribution of P13 differed significantly with that of P14. In case of Exponential distribution, the distribution of population P31 differed significantly with that of P32 and P33 with that of P34 (Table 3).

COMPARISONS AMONG SAMPLES, FOLLOWING UNIFORM DISTRIBUTION

The results of comparisons of sample distributions of (P11, P12) and (P13, P14), of sample size 15, by the selected α levels, are shown in Table 4.

At 5% α level, for the Uniform distribution samples of size 15, the Mann Whitney test could pick up correctly the expected difference, only in 7.9% of the comparisons while this percentage raised to 15.3% and 22.6% for α levels of 10% and 15%, respectively.

Table 3: Results of Comparisons between Different Population Distributions

Population	P11	P12	P13	P14	P31	P32	P33	P34
Mean	60.8	51.4	34.9	30.4	247.7	213.1	31.9	41.6
SD	33.94	28.97	22.55	17.53	257.58	252.36	17.81	27.5
Median	58.6	49.7	32.5	28.7	136.4	96.2	27.2	33.2
χ^2 Value	10.04		12.71		22.92		10.48	
df	4		4		5		4	
P-Value	< 0.05		< 0.01		< 0.001		< 0.05	

Table 4: Comparisons of Sample Distributions of (P11, P12) and (P13, P14) By Varying α levels and the Sample size of 15 - Mann Whitney test

Population Distribution	Populations	Sample size	Number of Comparisons	5%	10%	15%
Uniform	P11P12	15	1000	79	159	236
			%	7.9	15.9	23.6
	P13P14	15	1000	79	147	216
			%	7.9	14.7	21.6
POOLED		15	2000	158	306	452
			%	7.9	15.3	22.6

The results of comparisons of sample distributions of (P11, P12) and (P13, P14), of sample size 20, by the selected α levels, are shown in Table 5.

At 5% α level, for the Uniform distribution samples of size 20, the Mann Whitney test could pick up correctly the expected difference, only in 9.5% of the total sample comparisons. This percentage raised to 17.1% and 22.8% for α levels of 10% and 15%, respectively.

Table 5: Comparisons of Sample Distributions of (P11, P12) and (P13, P14) By Varying α levels and the Sample size of 20 - Mann Whitney test

Population Distribution	Populations	Sample size	Number of Comparisons	5%	10%	15%
Uniform	P11P12	20	1000	92	174	236
			%	9.2	17.4	23.6
	P13P14	20	1000	98	167	220
			%	9.8	16.7	22.0
POOLED		20	2000	190	341	456
			%	9.5	17.1	22.8

The results of comparisons of sample distributions of (P11, P12) and (P13, P14), of sample size 25, by the selected α levels, are shown in Table 6.

Table 6: Comparisons of Sample Distributions of (P11, P12) and (P13, P14) By Varying α levels and the Sample size of 25 - Mann Whitney test

Population Distribution	Populations	Sample size	Number of Comparisons	5%	10%	15%
Uniform	P11P12	25	1000	114	188	260
			%	11.4	18.8	26.0
	P13P14	25	1000	107	184	250
			%	10.7	18.4	25.0
POOLED	25	2000	221	372	510	
		%	11.1	18.6	25.5	

At 5% α level, for the Uniform distribution samples of size 25, the Mann Whitney test could pick up correctly the expected difference, only in 11.1% of the total sample comparisons while this percentage raised to 18.6% and 25.5% for α levels of 10% and 15%, respectively.

COMPARISONS AMONG SAMPLES, FOLLOWING EXPONENTIAL DISTRIBUTION

The results of comparisons of sample distributions of (P31, P32) and (P33, P34), of sample size 15, by the selected α levels, are shown in Table 7.

At 5% α level, for the Exponential distribution samples of size of 15, the Mann Whitney test could pick up correctly the expected difference, only in 11.6% of the total sample comparisons while this percentage raised to 19.0% and 26.4% for α levels of 10% and 15%, respectively.

Table 7: Comparisons of Sample Distributions of (P31, P32) and (P33, P34) By Varying α levels and the Sample size of 15 - Mann Whitney test

Population	Populations	Sample size	Number of Comparisons	5%	10%	15%
Exponential	P31P32	15	1000	121	192	263
			%	12.1	19.2	26.3
	P33P34	15	1000	111	187	264
			%	11.1	18.7	26.4
POOLED	15	2000	232	379	527	
		%	11.6	19.0	26.4	

The results of comparisons of sample distributions of (P31, P32) and (P33, P34), of sample size 20, by the selected α levels, are shown in Table 8.

At 5% α level, for the Exponential distribution samples of size 20, the Mann Whitney test could pick up correctly the expected difference, only in 11.4% of the total sample comparisons while this percentage raised to 20.2% and 26.4% for α levels of 10% and 15%, respectively.

The results of comparisons of sample distributions of (P31, P32) and (P33, P34), of sample size 25, by the selected α levels, are shown in Table 9.

Table 8: Comparisons of Sample Distributions of (P31, P32) and (P33, P34) By Varying α levels and the Sample size of 20 - Mann Whitney test

Population	Populations	Sample size	Number of Comparisons	5%	10%	15%
Exponential	P31P32	20	1000	112	192	256
			%	11.2	19.2	25.6
	P33P34	20	1000	115	211	272
			%	11.5	21.1	27.2
POOLED		20	2000	227	403	528
			%	11.4	20.2	26.4

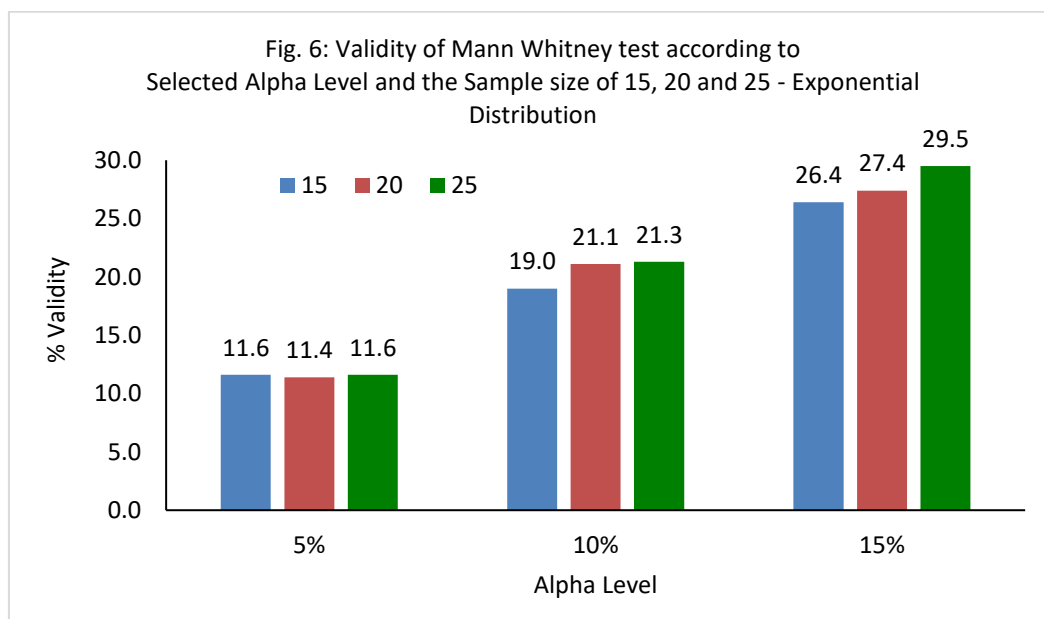
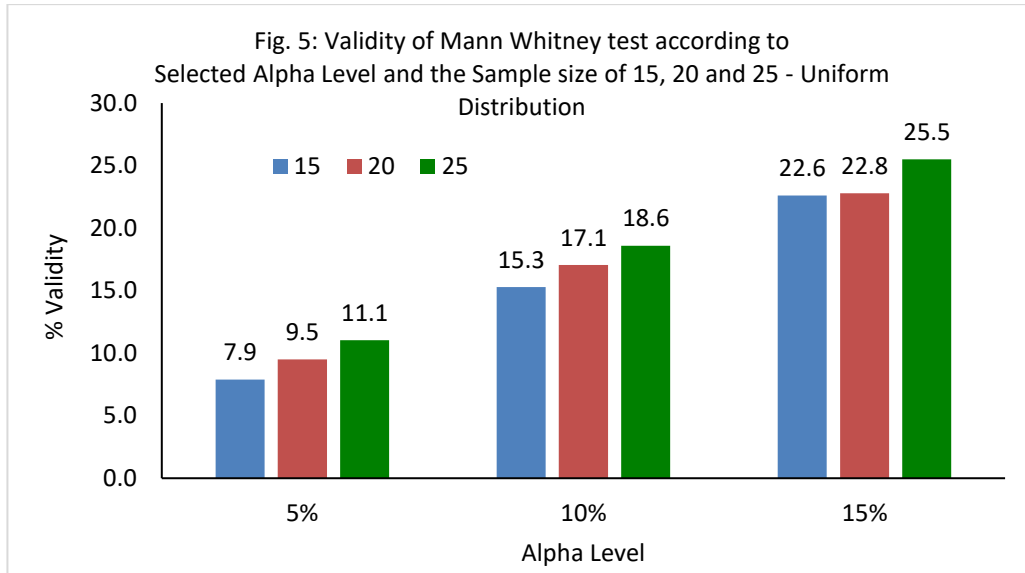
Table 9: Comparisons of Sample Distributions of (P31, P32) and (P33, P34) By Varying α levels and the Sample size 25 - Mann Whitney test

Population	Populations	Sample size	Number of Comparisons	5%	10%	15%
Exponential	P31P32	25	1000	106	184	273
			%	10.6	18.4	27.3
	P33P34	25	1000	126	242	319
			%	12.6	24.2	31.9
POOLED		25	2000	232	426	592
			%	11.6	21.3	29.6

At 5% α level, for the Exponential distribution samples of size 25, the Mann Whitney test could pick up correctly the expected difference, only in 11.6% of the total sample comparisons while this percentage raised to 21.3% and 29.6% for α levels of 10% and 15%, respectively.

VALIDITY OF MANN WHITNEY TEST AND THE SAMPLE SIZE

The validity of the Mann Whitney test according to selected α level and the sample size, for the Uniform Distribution, is shown in Fig. 5. At 5% α level, the validity of the test remained below 12% for all the three selected sample sizes. Again, at 15% α level, though the validity increased but remained below 26% for all the selected sample sizes.



The validity of the Mann Whitney test according to selected α levels and the sample sizes, for the Exponential distribution, is shown in Fig. 6. At 5% α level, the validity of the test remained below 12% for all the three selected sample sizes. Again, at 15% α level, the validity registered an increase but remained below 30% for all the selected sample sizes.

DISCUSSION

It is a basic rule to use parametric test whenever the data is following the normal distribution. For a non-normal data or for a skewed data, a non-parametric test is used. One of the popular beliefs

is that a non-parametric test can be used even for small samples (Gupta SC 2012, Lisa Sullivan 2017, Neil RS 2017, Francis SN 2021) as evident by the examples with 5-10 samples size used by them to demonstrate the application of the Mann Whitney test. In my recent publication, it was shown that for normal samples of size 10, t-test as well as Mann Whitney test have a low validity (Takiar 2023). In the current study, the samples of size 15, 20 and 25 are generated with different ranges (Uniform distribution) and rates (Exponential distribution) and when compared for possible significant differences between the distributions, a very few tests have shown correctly the expected significant differences between the distribution of samples. The validity is low for all the comparisons and is below 15% when $\alpha = 5\%$. Even on raising the α level to 15%, the validity did not rise much and remained below 30% which is equivalent to the power of the test. This suggests that for non-normal samples and sample size 25 and below, the use of Mann Whitney test with such a low power is not appropriate. So, the belief and the practice that a non-parametric test can be used for small samples even below 10 and for non-normal distributions, should be questioned and the test should be avoided for samples following the Uniform or the Exponential distribution, as shown in the current study. The suitability of the test for the samples of other non-normal distributions, need to be explored.

CONCLUSION

- At $\alpha = 5\%$, the Mann Whitney test showed a very low validity, below 15%.
- When α level is increased to 15%, even then the validity of the test remained below 30%.
- The test is not at all suitable when dealing with the samples of Uniform and Exponential distribution.
- In general, the statement that the non-parametric tests are suitable for any non-normal sample comparisons is not correct and should be used with due caution.
- The suitability of the test for the samples of other non-normal distributions, namely other than Uniform and Exponential distribution, need to be explored.

RECOMMENDATIONS

- For the use of any non-parametric test, consider $\alpha = 10\%$ or 15% , in order to increase the validity of the test.
- The Mann Whitney test should not be carried out for the distribution comparisons if the sample sizes under considerations are below 25.

REFERENCES

- [1]. Francis S.N. 2016. Nonparametric statistical tests for the continuous data: the basic concept and the practical use, Korean J Anesthesio. 69(1): Page 8–14. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4754273/>
- [2]. Gupta SC 2012: Non-parametric Methods, Fundamental of Statistics, Seventh Edition, Himalaya Publishing House; Page 26.1-26.2, 26.19-26.24.
- [3]. Gupta SC, Kapoor VK 2001: Fundamentals of Mathematical Statistics, Sultan Chand & Sons; Tenth revised Edition, Page 16.59-16.66.
- [4]. IBM Corp. (2015). IBM SPSS Statistics for Windows -Version 23.0, IBM Corp.
- [5]. Lisa Sullivan 2017. Non-parametric test – Boston University, School of Public Health

- https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric_print.html
- [6]. Microsoft Corporation, 2019. Microsoft Excel, Available at: <https://office.microsoft.com/excel>.
- [7]. Neil R.S. 2017. Non-Parametric Test-The Mann Whitney U Test
<https://www.sciencedirect.com/topics/medicine-and-dentistry/nonparametric-test>
- [8]. StatPlus 7.6.5.0 2021, AnalystSoft Inc.-Statistical analysis program. Version v7. See <https://www.analystsoft.com/en/>
- [9]. Takiar R (2023): 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.

Biography of corresponding author: 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 65 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 taught 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 7 research papers during 2021-23.