Review Article

Non–Parametric Test in Pharmaceutical Statistical Calculations

Syeda Bushra Fatima[1]*, Dr. K. Anupama [2]
1* Sultan –ul- Uloom College of Pharmacy, Department of Humanity, Telangana, India.
2 Sultan –ul- Uloom College of Pharmacy, Department of Pharmacology, Telangana, India.
***Corresponding author's E-mail:** syedafatima@sucp.ac.in

Received: 22-09-2022; **Revised:** 26-11-2022; **Accepted:** 05-12-2022; **Published on:** 15-12-2022.

ABSTRACT

Non parametric statistics, also known as distribution –free statistics, are methods of testing hypotheses when the nature of the distributions is not known, and we are not willing to accept the assumptions necessary for the application of the unusual statistical procedures. Nonparametric tests are the statistical methods based on signs and ranks. In this article, the concepts and use of nonparametric tests is described under pharmaceutical statistics.

Keywords: Non Parameter, Data interpretation, Non-parametric statistics, Statistical data analysis.

QUICK RESPONSE CODE →

DOI:
10.47583/ijpsrr.2022.v77i02.009

DOI link: http://dx.doi.org/10.47583/ijpsrr.2022.v77i02.009

INTRODUCTION

The nonparametric method is a type of statistic that does not make any assumptions about the characteristics of the sample (parameters) or whether the observed data is quantitative or qualitative[2].In Non-Parametric tests, we don't make any assumption about the parameters for the given population or the population we are studying. These tests don't depend on the population. Hence, there is no fixed set of parameters is available, and also there is no distribution (normal distribution, etc.) of any kind is available for use. This is also the reason that nonparametric tests are also called as distribution-free tests. Most of the nonparametric tests are very easy to apply and to understand also i.e. the complexity is very low[3]. Non parametric test are based on simple assumptions i.e. sample observation are independent of each other, variables are continuous in nature and probability distribution function is continuous.[4]

Parametric tests are used when the information about the population parameters is completely known whereas non-parametric tests are used when there is no or few information available about the population parameters i.e. parametric test assumes that the data is normally distributed. However, non-parametric tests make no assumptions about the distribution of data. Nonparametric statistical tests can be useful alternative to parametric statistical tests when the test assumptions about the data distribution are not met.[6] Non parametric tests are most effectively used for data which consist of only classified (nominal) variables or ranked variables which are considered to have underlying continuous distribution. Data retain from continuous distributions are particularly manageable to non-parametric methods, but more complex designs in which interactions and other ANOVA components are present cannot be simply analyzed with these techniques, particularly when sample sizes are small.[1] Most of the Non parametric methods for data that are not categorical use ranking procedures. Observations in various treatment groups are ranked according to specific procedures, and the ranks that replace the raw data are analyzed. These analyses use simpler statistical computations than the corresponding parametric analyses. Transformation to ranks results in simple whole or fractional numbers of relatively small magnitude.[1]

Categories/Types of Non-Parametric Test

Nonparametric methods are classified according to their function, such as two-sample tests and so on. Since populations do not always meet the assumptions underlying parametric tests, we frequently need probable procedures whose validity is not dependent on rigid assumptions. Nonparametric statistical procedures fill this need in many instances, since they are valid under very general assumptions[6]

Sign test: This test is use to estimate the median of a population and compare it to a reference value or target value or critical value.

Wilcoxon signed rank test: In this test, one can estimate the population median and compare it to a reference/target value/critical value. However, the test assumes your data comes from a symmetric distribution (like the Cauchy distribution or uniform distribution).[7]

Wilcoxon signed rank sum test: Tests the equality of the distribution of the two treatments.[1]

Friedman test: Test for differences between groups with ordinal dependent variables. It can also be used for continuous data if the one-way ANOVA with repeated measures is inappropriate (i.e. some assumption has been violated).

Kruskal-Wallis test: This test is use instead of a one-way ANOVA to find out if two or more medians are different. Ranks of the data points are used for the calculations, rather than the data points by subjectively.

The Mann-Kendall Trend Test looks for trends in time-series data.[7]

NONPARAMETRIC TEST	PARAMETRIC ALTERNATIVE
1-sample sign test	One-sample Z-test, one sample t-test
1-sample Wilcoxon Signed Rank Test	One Sample Z-test, one sample t-test
Friedman test	Two-way ANOVA
Kruskal-Wallis test	One-way ANOVA
Mann-Whitney test	Independent samples t-test[7]

Advantages:

1. Deliver accurate conclusion even when the sample size is small.
2. It's more powerful than parametric tests when assumptions of normality have been violated.
3. Suitable for all data types, such as nominal, ordinal, interval or the data which has outliers.[5]

Disadvantages:

May Waste Information:- Researcher may waste information when parametric procedures are more appropriate to use. If the assumptions of the parametric methods can be met, it is generally more efficient to use them.

Difficult to compute by hand for large samples:- For large sample sizes, data manipulations tend to become more strenuous, unless computer software is available.

Tables not widely available: Often special tables of critical values are needed for the test statistic, and these values cannot always be generated by computer software. On the other hand, the critical values for the parametric tests are readily available and generally easy to incorporate in computer programs[6]

If exists any parametric test for a data then using non-parametric test could be a terrible blunder.

The critical value tables for non-parametric tests are not included in many computer software packages so these tests require more manual calculations.[5]

DISCUSSIONS

Sign Test[13]

Sign test is one of the simplest non- parametric tests. It is a test of the equality of the medians of two comparative groups[1]. It is based on the direction of the plus or minus signs of observations in a sample and not on their numerical magnitudes.[5] It is used for paired data with an underlying continuous distribution, and is be applied to ranked or higher level data such as continuous interval and ratio-type data. The pairs are matched, and differences of the measurements for each pair tabulated. The differences are then classified only with regard to the sign of the difference. i.e count the number of times one treatment has a higher value than the other. If the test shows significance then two comparative populations have different medians at α level of significance. The statistical test is based on binomial distribution.[1] The sign test is used in the following way:

Number of plus (+) signs

Number of minus (-) signs

Number of 0's which cannot be included either as positive or negative.[8]

When applying two treatments to the same person, there are two possible outcomes: either treatment A is favoured or treatment B is favoured.[1]

Question:-

Compare the "time to peak" plasma level for two oral formulations of the same drug. Data is analysed using more sensitive non parametric test or a t- test for paired data (or ANOVA for a crossover design). Values are obtained by administering both drugs to each of 12 persons on two different occasions. Although the data would ordinarily result from a crossover design, and ANOVA techniques might be more appropriate, for current purposes, consider an example where treatments have assigned in random order. Assume the 'order 'effects will not analyse and no carryover effects are present.

Table 1: Paired Data obtained from the bioavailability experiment: Time to peak plasma concentration Time to peak (hr.)

Subject	A	B	Difference: B-A
1	2.5	3.5	+1
2	3.0	4.0	+1
3	1.25	2.5	+1.25
4	1.75	2.0	+0.25
5	3.5	3.5	+0
6	2.5	4.0	+1.5
7	1.75	1.5	-0.25
8	2.25	2.5	+0.25
9	3.5	3.0	-0.5
10	2.5	3.0	+0.5
11	2.0	3.5	+1.5
12	3.5	4.0	+0.5

Tabulation of the differences (B-A) results in Nine positive signs and two negative signs. One subject showed no difference between treatments A and B. Table value (IV.12) 10 of 11 positive (or negative) signs are needed to obtain significance at the 5% level. According to sign test, the difference just misses significance, although product B appears to take longer time to peak than product A.

If differences are assumed to normal distribution, the paired t-test would be a more sensitive test than the sign test. For specific example one could not predict that the t-test would result in a 'more significant' difference; but on average the t-test will be more discriminating. Above example conclude that the t-test results in a highly significant difference between the two formulations i.e. t=3.02.[1]

Wilcoxon's signed rank test[14]

WILCOXON MATCHED PAIRS SIGNED RANK TEST (WILCOXON SIGNED RANK TEST) is also called the Wilcoxon matched pairs test or the Wilcoxon signed rank test. It is very appropriate for a repeated measure design where the same subjects are evaluated under two different conditions. It is the nonparametric equivalent of the parametric paired t-test. This is not the same as the Wilcoxon rank sum test, which compares two non-paired groups and is equivalent to the parametric unpaired t-test. The Wilcoxon signed rank is more powerful than the sign test. This statistic differs from the sign test in that it considers the magnitude of the difference while the sign test does not. It uses more information from the sets of scores than the simple sign test. Because it uses more information it is considered to be more precise than the sign test.[9]

Wilcoxon test is based on the assumption that the distributions of the comparative treatments are symmetrical. We test the equality of the means or medians; the mean and median are equal in a symmetrical distribution.[1]

Initial calculations are the same as the sign test. Firstly take the differences between the treatment pairs as above example. When the values for a treatment pair are equal (difference of zero), a tie, and these data are discarded for purposes of the test. As in sign test, zero difference does not contribute information regarding the differentiation of treatments in the Wilcoxon signed rank test.

Table 2: Data from above example: Ranking Differences without regard to Sign test for Wilcoxon signed Rank test

Subject	Value	Rank	Assigned rank	Assigned rank with sign	Ranks with positive signs	Ranks with negative signs
7	-0.25	1	2	-2		2
4	0.25	2	2	2	2	
8	0.25	3	2	2	2	
9	-0.5	4	5	-5		5
10	0.5	5	5	5	5	
12	0.5	6	5	5	5	
1	1.0	7	7.5	7.5	7.5	
2	1.0	8	7.5	7.5	7.5	
3	1.25	9	9	9	9	
6	1.5	10	10.5	10.5	10.5	
11	1.5	11	10.5	10.5	10.5	
					Sum=59	Sum=7

Differences of united pairs are ranked in order of magnitude, ignoring sign. The data under Table -1 is the comparison of time peak plasma concentration for two formulations A and B, and ranking of absolute values of the difference is shown Table-2. In above example the means are significantly different. The table shows that rank sum of 10 or less for smaller rank sum is significant at 5% level for N=11 i.e. the smaller sum rank is 7 so the difference is significant at 0.05 level (p~ 0.02).

For larger sample size normal approximation is $Z=\frac{|R-N(N+1)/4|}{\sqrt{[N(N+\frac{1}{2})(N+1)]/12}}$ → (eq:-1)

Where R=sum of ranks; N=sample size

For above example; $Z=\frac{|59-11(12)/4|}{\sqrt{[11(11.5)(12)]/12}}=2.31$

From table P=0.02, which is very close to exact probability, if data are normally distributed.[1]

The sign test is limited in that it cannot reflect the degree of change between paired scores. Wilcoxon's signed rank test has more statistical power than the sign test because it not only considers the direction of the change but also ranks the degree of change between the paired scores, providing more information for the analysis.[11]

Wilcoxon Signed Rank Sum Test [15]

Wilcoxon signed rank sum test is analogous to two sample t-test and is stronger than the sign test for paired observations. It compares two samples that drawn from two independent populations. It evaluates the null hypothesis that the two distributions are identical against the alternative hypothesis that the two distributions differ only with respect to the median.[10] The calculations for Wilcoxon rank sum test are similar to those for the signed rank test.

Table 3: Results of a Dissolution Test using the USP Dissolution apparatus and a modification: Amount Dissolved in 30 minutes

USP apparatus		Modified apparatus	
Amount dissolved	**Rank**	**Amount dissolved**	**Rank**
53	3	58	11
61	14	55	5.5
57	9	67	21
50	1	62	15.5
63	17	55	5.5
62	15.5	64	18.5
54	4	66	20
52	2	59	12.5
59	12.5	68	22
57	9	57	9
64	18.5	69	23
Sum of Ranks	105.5	56	7
		Sum of Ranks	170.5

Above table -3 shows the comparison of USP Dissolution apparatus1 and a modified apparatus with 30 min. The objective is to compare the performance of the two pieces of apparatus. Twelve individual tablets were used for each "treatment" (apparatus). The amount of drug dissolved in 30min was determined for each tablet. One tablet assay, determined in the original apparatus, is not included in the results because of an overt error during the assay procedure for this tablet.

For moderate sized samples, the statistical test for equality of the distribution means may be approximated using the normal distribution. The normal approximation is

$$Z = \frac{|T - N1(N1+N2+1)/2|}{\sqrt{[N1N2(N1+N2+1)]/12}} \quad \rightarrow \text{(eq:-2)}$$

Where N_1= smaller sample size

N_2=larger sample size

T=sum of ranks for smaller sample size

If Z ≥ 1.96 the two treatments can be significantly different at 5% level (Two sided test). Then for our example

$$Z = \frac{|105.5 - 11(11+12+1)/2|}{\sqrt{[11(12)(11+12+1)]/12}} = \frac{26.5}{16.25} = 1.63$$

A value of Z=1.63 is not large enough to show significance in a two – sided test at the 5% level (p=0.11). Therefore, these data do not provide sufficient evidence to show that two different pieces of apparatus give different dissolution results. The t- test is more efficient than the non-parametric rank sum test if the assumptions for the t-test are valid. Similar to the signed rank test, the Wilcoxon rank sum test is very efficient, approximately 95% compared to the corresponding t-test i.e. t= 1.84 with 21 degree of freedom. The non-parametric tests are useful in experiments where the data consist of values derived from a rating scale with an underlying continuous distribution.[1]

Friedman Test (Two –way ANOVA) [16]

The Friedman test is a non-parametric test applied to data which is, at least ranked and which is in the form of a two-way ANOVA design (randomized blocks).[1] This test is applied when data is in the form of a twoway ANOVA design and data should be suitable for ranking and also should be from more than two groups. It is for use with K repeated or correlated measures where measurement is at least ordinal. The null hypothesis states that all k samples are drawn from the same population or from populations with equal median.it is the significance test for more than two dependent variables.[9]

Table 4: Average Hardness of 10 Tablets for five different tablet formulations prepared on four presses[a]

Tablet formulation	Tablet press			
	A	**B**	**C**	**D**
1	7.5(4)	6.9(1)	7.3(3)	7.0(2)
2	8.2(3)	8.0(2)	8.5(4)	7.9(1)
3	7.3(1)	7.9(3)	8.0(4)	7.6(2)
4	6.6(3)	6.5(2)	7.1(4)	6.4(1)
5	7.5(3)	6.8(2)	7.6(4)	6.7(1)
R_i	14	10	19	7

[a]Parenthetical values are the within-tablet-press ranks.

Data above describes the results of a validation experiment to test the performance of four tablet presses, with regard to tablet hardness. Average hardness of 10 tablets was computed for five different tablet products manufactured on four presses. Tablets are a random selection of five typical tablet products. The presses were identically set for same pressure for tablet formulation. The parenthetical values in table 4 are ranks of the average hardness for each formulation over the four presses.

If one of the presses consistently had the highest (or lowest) rank, one would conclude that the press (treatment)produced harder(or less hard) tablets than the other presses. In table 4, tablet press C had the highest hardness value for all formulations with the exception of formulation 1, where it had the next-to-largest value. The

test of significance is an objective assessment of whether or not the data of given table provide sufficient evidence to say that tablet C is, indeed, producing harder tablets than the other presses.

Chi –square distribution is used for sample sizes sufficiently large to approximate the test of significance. The chi square test is

$$X^2_{c-1} = \frac{12}{rc(c+1)} (\Sigma R_i^2 - 3r(c+1)) \rightarrow \text{(eq :- 3)}$$

where $X^2_{c-1} = the\ x^2 static\ with\ c-1\ degree\ of\ freedom$

$$r = number\ of\ rows(blocks)$$

$$c = number\ of\ columns\ (treatments)$$

R_i =sum of ranks in i^{th} group (column)

According to Table -4 we have, the chi- square statistic has 3 degree of freedom.

$$X^2_3 = \frac{12}{(5)(4)(4+1)} (14^2 + 10^2 + 19^2 + 7^2) - 3(5)(5) = 9.72$$

A chi – square value of 7.81 or larger is needed for significance at 5% level. We conclude that at least two of the tablet presses differ with regard to tablet hardness. Observation shows that tablet press C produces harder tablets than those produced by the other presses. The tabulated value shows a difference of 11 is required for significance ($p < 0.05$) for individual comparisons between pairs of means for 4 treatments (k=4) and 5 rows (n=5). Henceforth, press C produces significantly harder tablets than press D with sum of ranks of 19 and 7 respectively.

Kruskal-Wallis H test (one- way ANOVA)[17]

The Kruskal-Wallis H test is a rank based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. For example in clinical trials, it will be used to test assess differences in albumin levels in adults different diets with different amounts of protein.[12]

It is an extension of the rank sum test to more than two treatments, and is a test of the location of distributions. Significance: differences can be interpreted as meaning that the averages of at least two of the comparative treatments are different. The computations and analysis will be illustrated using an experiment in which data were obtained from a preclinical experiment in which rats, injected with two doses of an experimental compound and a control (a known sedative), were observed for sedation. The time for the animals to fall asleep after injection was recorded. If an animal did not fall asleep within 10min of the drug injection, the time to sleep was arbitrarily assigned a value of 15min.table 5 shows the experimental results.one data point was lost from the control group because of an illegible recording, obliterated in the laboratory notebook.

Table 5: "Time to sleep" for a control and Two Doses of an Experimental compound (min)

Control	Rank	Low dose	Rank	High dose	Rank
8	22	10	26	3	10
1	3.5	5	13	4	12
9	24.5	8	22	8	22
		6	15	1	3.5
9	24.5	7	18.5	1	3.5
6	15	7	18.5	3	10
3	10	15	28	1	3.5
15	28	1	3.5	6	15
1	3.5	15	28	2	7.5
7	18.5	7	18.5	2	7.5
sum of ranks	149.5		191.0		94.5

The analysis for treatment differences is not dependent on equal numbers of observations per group, although, as in most experiments, equal sample sizes are most desirable (optimal). Analysis consists of first combining all of the data then obtains the ranks then the observations are reclassified into original groups.

Test statistic for kruskal-wallis test, as described below is approximately distributed as chi- square with k – 1 d.f, where k= number of treatments (groups). For small sample sizes, tables to determine the treatment rank sums needed for significance. Chi square approximation is good if the number of observations in each group is greater than 5.the computation of chi-square statistic follows:

$$X^2_{k-1} = \frac{12}{N(N+1)} \left(\Sigma \frac{R_i^2}{n_i}\right) - 3(N+1) \rightarrow \text{(eq:-4)}$$

Where N= total number of observations in all groups combined

R_i= sum of ranks in i^{th} group

n_i= number of observations in i^{th} group

k= number of groups

For Table-5, N=29, R_1=149.5, R_2=191, R_3=94.5, n_1= 9, n_2= 10,n_3=10,k=3

we have,

$$X^2_2 = \frac{12}{29(30)} \left(\frac{149.5^2}{9} + \frac{191^2}{10} + \frac{94.5^2}{10}\right) - 3(29+1) = 6.89$$

The chi-square value with 2 d.f must be equal to or greater than 5.99 to be significant at 5% level. Hence the average "time to sleep" differs for at least two of the three treatment groups (control, high dose and low dose) at the 5% level of significance.

A correction for ties can be used which increases the value of chi-square. Hence, if null hypothesis is rejected (significant treatment differences), the correction only increases the degree of significance. If chi-square just

misses significance, the correction result in statistically significant differences. The correction as follows

$$\text{Correction} = \frac{x^2}{1-\Sigma(t_i^2-t_i)/(N^3-N)} \rightarrow \text{(eq:-5)}$$

Where t_i = number of tied observations in group "i"
N= total number of observations

Hence, $\frac{6.89}{1-378/(29^3-29)} = \frac{6.89}{0.984} = 7.00$

The correction for ties is usually very small. The correction for above example is that does not change the conclusion of significant differences among treatment means.[1]

CONCLUSION

The aim of this article is to emphasis on Non-parametric statistics under pharmaceutics which is without regard to any primary distribution. It is convenient alternative to parametric statistical tests when the test assumptions about the data distribution are not met. The main purpose of nonparametric tests is where the methods of statistical analysis do not require a distribution to meet the required assumptions to be analysed**.**

In clinical practice, the researchers often assess whether the outcome variable is overall normally distributed and use a nonparametric test when it is not.

REFERENCES

1. Bolton S. Pharmaceutical statistics: Practical and clinical applications. 2nd ed. New York: Marcel Dekker, Inc; 1990

2. Hayes A. Non parametric method [Internet]. New York: Dotdash Meredith; 2021 [updated 2021]. Available from: https://www.investopedia.com/terms/n/nonparametric-method.asp

3. Goyal A. Hypothesis Testing - Parametric and Non-Parametric Tests in Statistics [Internet]. India: Analytics Vidhya; 2021 [updated 2021]. Available from: https://www.analyticsvidhya.com/blog/2021/06/hypothesis-testing-parametric-and-non-parametric-tests-in-statistics/

4. Surya RK. Biostatistics for health and life sciences. Mumbai: Himalaya Publishing house; 2010.

5. Av content team. A guide to conduct analysis using Non-parametric statistical tests [Internet]. India: Analytics Vidhya; 2017 [updated 2020]. Available from: https://www.analyticsvidhya.com/blog/2017/11/a-guide-to-conduct-analysis-using-non-parametric-tests/

6. Hesse CA, Ofosu JB, Nortey EN. Introduction to nonparametric statistical methods. Accra: Akrong Publications Ltd; 2017.

7. Glen S. Non parametric data and tests (Distribution free tests) [Internet]. c2022. Available from:

https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/parametric-and-non-parametric-data/

8. Wisdom IT services India Pvt. Ltd. Important non parametric or distribution free tests - Research Methodology [Internet]. c2020. Available from: https://www.wisdomjobs.com/e-university/research-methodology-tutorial-355/important-nonparametric-or-distribution-free-tests-11568.html

9. Khan IA, Khanum A, Khan S. Fundamentals of Biostatistics. 5th ed. Hyderabad: Ukaaz Publication; 2018.

10. Scheff SW. Fundamental statistical principles for the neurobiologist: A survival guide. United States of America: Academic Press; 2016.

11. Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. Korean J Anesthesiol. 2016 Feb 1;69(1):8-14.

12. Goswami A. Utilization of Parametric and Non-Parametric Statistical Test in Clinical Trials. International Journal of Science and Research. 2016; 5(10): 1445-1449.

13. Sign Test [Internet]. 2005 [updated 2022]. Available from: https://en.m.wikipedia.org/wiki/Sign_test

14. Schwaid G. Epidemiology and biostatistics. Board Review in Preventive Medicine and Public Health. Academic Press; 2017:79-185.

15. Pharmaceutical Biostatistics [Internet]. 2019. Available from: https://issuu.com/pharmaceuticalbiostatistics/docs/4-wilcoxon_signed_rank_test

16. Glen S. Friedman's Test/ Two Way Analysis of Variance by ranks [Internet]. c2022. Available from: https://www.statisticshowto.com/friedmans-test/

17. [Internet]. c2022. Available from: https://www.statstest.com/kruskal-wallis-one-way-anova/

Source of Support: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.