International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 5 Issue 9, September 2018

ISSN: 2394-4404

Nonparametric Tests

Isah Mu'azu

Department of Mathematics Sokoto State University

Abstract: ***This paper discusses the three of the most common Nonparametric tests, the Kruskal Wallis, the Wilcoxon signed rank and Friedman test. Example for performing Kruskal wallis test were demonstrated. Post-hoc test in this case the Bonnferroni test was also explained.***

I. INTRODUCTION

"It is fashionable to claim that nonparametric methods were first used when J. Arbuthnot (1710) found that in each year from 1929 to 1710 the number of males christened in London exceeded the number of females. He regarded this as evidence that the probabilities of any birth being male or female where not exactly equal, a discrepancy Arbuthnot attributed to Devine providence. A sign test is appropriate for his data" Sprent (1989. P: 5).

The nonparametric test Statistics is one which requires few assumptions to be met," The nonparametric tests do not assume the data is normal; rather they can be used on a much wider experiments. We sometimes want to make inferences that have nothing to do with parameters; or we may have data in a form that makes; say, normal theory tests inappropriate; we may not have precise measurement data, but only the rank order of observations. In general, these methods are applicable to estimation or hypothesis-testing problems when the populations distributions need only be specified in broad terms, e.g. as being continuous, symmetric, identical, differing to specific families the normal, uniform, exponential, etc. logically, the term distribution-free may then be more appropriate than nonparametric, but the latter term is well established in popular usages" Sprent (1989. P: 2-3).

A. DEFINITION

According to Investopedia,

"Statistical method wherein the data is not required to fit a normal distribution. Nonparametric statistics uses data that is often ordinal, meaning it does not rely on numbers, but rather a ranking or order of sorts. For example, a survey conveying consumer preferences ranging from like to dislike would be considered ordinal data.

Nonparametric statistics have gained appreciation due to their ease of use. As the need for parameters is relieved, the data becomes more applicable to a larger variety of tests. This type of statistics can be used without the mean, sample size, standard deviation, or estimation of any other related parameters when none of that information is available".

B. WHEN IS NONPARAMETRIC TEST APPROPRIATE TO USE

It can sometimes be difficult to assess whether a continuous outcomes follows normal distribution or not, and thus whether parametric or nonparametric is appropriate. There are several Statistical tests that can be used to assess whether data are likely from a normal distribution. The most popular are Kolmogorov-Smirnov test; the Anderson Darling test and the Shapiro wilk test see Taylor et.al (2011).

C. DISTINCTION OF NONPARAMETRIC TEST

There are several advantages attached to the use of nonparametric methods according to "mann (2001.P: 618-619) they are easier to use and understand; they can be applied to situations in which parametric tests cannot be used and they do not require that the population being sampled is normally distributed. However, a major problem with nonparametric tests is that they are less efficient than parametric. The sample size must be larger for a non-parametric to have the same probability of committing the two types of errors".

"Statisticians have shown that nonparametric tests are often very nearly as good as parametric tests even in the exact case for which the parametric tests are more designed. To illustrate, suppose two populations distribution are normal in shape and have the same standard deviation but different means. Then, the wilcoxon test is very nearly as powerful in detecting this difference as the t test, even though it uses only the ranks of the observations" Agresti and Franklin (2009.p:754).

Although nonparametric tests that can be applied are numerous, in this paper will focus on the following techniques:

Kruskal-Wallis Test
Wilcoxon's signed Rank Test
Friedman Test

II. KRUSKAL-WALLIS TEST

OVERVIEW

This statistical technique, often used in place of one way analysis of variance, if the data in question does not satisfy normality assumption and the sample size is relatively small." A popular nonparametric test to compare outcomes among more than two independent groups is the kruskal Wallis test. The kruskal wallis test is used to compare medians among K comparison groups (K>2) and is sometimes described as an ANOVA with the data replaced by their ranks. "When the Kruskal-Wallis test leads to significant results, then at least one of the samples is different from the other samples. The test does not identify where the differences occur or how many differences actually occur. It is an extension of the Mann–Whitney U test to 3 or more groups. The Mann-Whitney would help to analyze the specific sample pairs for significant differences. The Kruskal-Wallis ANOVA is useful as a general nonparametric test for comparing two or more independent samples. It can be used to test whether such samples come from the same distribution. They are powerful alternatives to the one-way analysis of variance.

The Kruskal-Wallis ANOVA uses the sum of difference between mean ranks of these samples as the statistic. The actual statistic of Mood's median test only relates to the number of larger or smaller than the median value but not their actual distance from the median, so it is not as effective as Kruskal-Wlalis ANOVA.

"As an example, researchers want to know whether the enhanced eyesight of young patients, who use three different therapies, to enhance their eyesight, comes from the same distribution. Thirty students' enhance eyesight, after adopting these three therapies, was recorded. Following table is an example of single- factor example:

Treat ment 1	y_{11}	y_{12}	y_{13}	$\hat{Y} = (y_{11} + y_{12} + y_{13})/3$
Treat ment 2	y_{21}	y_{22}	y_{23}	$\hat{Y} = (y_{21} + y_{22} + y_{23})/3$
Treat ment 2	y_{31}	y_{32}	y_{33}	$\hat{Y} = (y_{31} + y_{32} + y_{33})/3$
				$\hat{Y} = (\hat{Y}_1 + \hat{Y}_2 + \hat{Y}_3)/3$

Table 1

Let Yij be the random variable representing jth data point of ith treatment.

$$Yij = \mu + \tau_i + \in ij$$

Where μ so-call overall mean is the mean over whole sample, τ_i called treatment effect denotes the parameter of ith treatment and $\in ij$ denotes the random error.

Basically of ANOVA focuses on characteristics relating to deviation, variability, sum of squares, mean squares etc. A typical approach of ANOVA is to test whether k treatment means $\mu_1, \mu_2, \ldots, \mu_k$ are equal; it means that we test the following hypotheses:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_1 : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$$

If H_0 is true, treatments have no effect on whole sample. Let Yij be the instance of random variable Yij. Let Yi. Be Ŷi. Y and Ŷ be the sum of observations of treatment i,the average of observation of treatment i, the sum of whole observations and the average of whole observations.

$$y_i = \sum_{j=1}^{ni} yij,\quad \hat{Y} = \frac{1}{ni} \sum_{j=1}^{ni} yij,\quad y = \sum_{i=1}^{k} \sum_{j=1}^{ni} yij$$
$$,\hat{Y} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{ni} yij.$$

Where K is the number of treatments, ni is the number of observations under treatment and *N = n1 + n2 +... + nk* is the total number of observations.

Let $SS_{Treatment}$ and SSE, be the total sum of squares, treatment sum of squares and error sum of squares. Please pay attention to $SS_{Treatment}$ and *SSE* because they are main research objects in ANOVA.

We have:

$$SS_T = \sum_{i=1}^{k} \sum_{j=1}^{ni} (Y_{ij} - \hat{Y})^2$$
$$SS_{Treatment} = \sum_{i=1}^{k} n_i (\hat{Y}_i - \hat{Y})^2$$
$$SS_E = \sum_{i=1}^{k} \sum_{j=1}^{ni} (Y_{ij} - \hat{Y})^2$$

Following is the sum of squares identity:

$$SS_T = SS_{Treatment} + SS_E$$

Treatment sum squares $SS_{Treatment}$ is very important because it reflects treatment effect effects τ_i (S) and means μ (S). the expected values of treatment sum of squares and error sum of squares are computed as below:

$$E(SS_{Treatment}) = (K-1)\partial^2 + \sum_{i=1}^{k} n_i \tau_i^2$$
$$E(SS_E) = (N-K)\partial^2$$

SS_T and $SS_{Treatment}$ and SS_E have N-1 and N-K degrees of freedom, respectively because there are N observations over whole sample and treatments .So SS_E has N-K=(N-1)-(K-1) due to $SS_T = SS_{Treatment} + SS_E$. Based on degrees of freedom, treatment mean square $MS_{Treatment}$ and error mean square MS_E is determined as below:

$$MS_{Treatment} = \frac{MS_{Treatment}}{K-1}$$

$$MS_E = \frac{SS_E}{N-K}$$

If null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = 0$ is true, $MS_{Treatment}$ is an unbiased estimate of variance ∂^2 due to $E(SS_{Treatment}) = \frac{1}{N-K} E(SS_{Treatment}) = \partial^2$

$+\frac{1}{N-K}\sum_{i=1}^{k} n_i \quad \tau_i^2 = \partial^2 + \frac{1}{N-K}\sum_{i=1}^{k} n_i \quad 0^2 = \partial^2$."Moreover MS_E is always an unbiased estimate of variance ∂^2 due to E (MS_E) = $\frac{1}{N-K}E(SS_E)$ = ∂^2. So $MS_{Treatment}$ and MS_E conform chi-square distribution and the ratio of $MS_{Treatment}$ to MS_E conforms F- distribution with K-1 and n(k-1) degrees of freedom:

$$F_0 = \frac{MS_{Treatment}}{MS_E} \sim F_{K-1,\ N-K}$$

Hypothesis $H_0 : \tau_1 \quad \tau_2 = \cdots = \mu_k = 0$ is rejected if the ratio $F_0 >$ $f\alpha, K-1, n(k-1)$ where $f\alpha, K-1, n(k-1)$

Is the 100α percentage point of F-distribution with K-1 and N-K degrees of freedom". Maurya et.al (2013. P.34-36).

Parametric ANOVA with normality assumption is treated, now nonparametric test the Kruskal-Wallis test. As it has been said early that it does not require the distribution to be normal but assumes population variance among groups are equal and applies to the ranks not original data (Vile, J. 2013 P.53-60).

The hypotheses is as follows:

H_0: The K population medians are equal

H_1: The k population medians are not all equal.

The procedure for the test involves pooling the observations from the K samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to N, where N = $n_1+n_2+ ...+ n_k$.

To illustrate the procedure, consider the following example.

A clinical study is designed to assess differences in albumin levels in adults following diets with different amounts of protein. Low protein diets are often prescribed for patients with kidney failure. Albumin is the most abundant protein in blood, and its concentration in the serum is measured in grams per deciliter (g/dL). Clinically, serum albumin concentrations are also used to assess whether patients get sufficient protein in their diets. Three diets are compared, ranging from 5% to 15% protein, and the 15% protein diet represents a typical American diet. The albumin levels of participants following each diet are shown below.

5% Protein	10% Protein	15% Protein
3.1	3.8	4.0
2.6	4.1	5.5
2.9	2.9	5.0
	3.4	4.8
	4.2	

Table 2

Is there any difference in serum albumin levels among subjects on the three different diets? For reference, normal albumin levels are generally between 3.4 and 5.4 g/dL. By inspection, it appears that participants following the 15% protein diet have higher albumin levels than those following the 5% protein diet. The issue is whether this observed difference is statistically significant.

In this example, the outcome is continuous, but the sample sizes are small and not equal across comparison groups (n_1=3, n_2=5, n_3=4). Thus, a nonparametric test is appropriate. The hypotheses to be tested are given below, and we will us a 5% level of significance.

H_0: The three population medians are equal versus

H_1: The three population medians are not all equal

To conduct the test we first order the data in the combined total sample of 12 subjects from smallest to largest. We also need to keep track of the group assignments in the total sample.

			Total Sample (Ordered Smallest to Largest)			Ranks		
5% Protein	10% Protein	15% Protein	5% Protein	10% Protein	15% Protein	5% Protein	10% Protein	15% Protein
3.1	3.8	4.0	2.6			1		
2.6	4.1	5.5	2.9	2.9		2.5	2.5	
2.9	2.9	5.0	3.1			4		
	3.4	4.8		3.4			5	
	4.2			3.8			6	
					4.0			7
				4.1			8	
				4.2			9	
					4.8			10
					5.0			11
					5.5			12

Table 3

Notice that the lower ranks (e.g., 1, 2.5, 4) are assigned to the 5% protein diet group while the higher ranks (e.g., 10, 11 and 12) are assigned to the 15% protein diet group. Again, the goal of the test is to determine whether the observed data support a difference in the three population medians. in the parametric tests, the hypothesis testing, when comparing means among more than two groups we analyze the difference among the sample means (mean square between groups) relative to their within group variability and summarize the sample information in a test statistic (F statistic). In the Kruskal Wallis test we again summarize the sample information in a test statistic based on the ranks.

Test Statistic for the Kruskal Wallis Test

The test statistic for the Kruskal Wallis test is denoted H and is defined as follows:

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^{\ 2}}{n_j} \right) - 3(N+1)$$

Where k=the number of comparison groups, N= the total sample size, n_j is the sample size in the j^{th} group and R_j is the sum of the ranks in the j^{th} group.

In this example R_1 = 7.5, R_2 = 30.5, and R_3 = 40. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 12(13)/2=78 which is equal to 7.5+30.5+40 = 78. The H statistic for this example is computed as follows:

$$H = \frac{12}{12(13+1)}\left[\frac{7.5^2}{3} + \frac{30.5^2}{5} + \frac{40^2}{4}\right] - 3(12+1) = 7.52$$

We must now determine whether the observed test statistic H supports the null or research hypothesis. Once again, this is done by establishing a critical value of H. If the observed value of H is greater than or equal to the critical value, we reject H_0 in favor of H_1; if the observed value of H is less than the critical value we do not reject H_0. The critical value of H can be found in the table below.

ISSN: 2394-4404

To determine the appropriate critical value we need sample sizes (n_1=3, n_2=5 and n_3=4) and our level of significance (α=0.05). For this example the critical value is 5.656, thus we reject H_0 because 7.52 ≥ 5.656, and we conclude that there is a difference in median albumin levels among the three different diets.

Our Statistical analyses are based mainly on probabilities on the view of the above we draw conclusions on the basis of the result available to us to a situation under consideration. "A statistics can never establish the truth of a hypothesis with 100 percent certainty. Typically, the hypothesis is specified in the form of a "null hypothesis," i.e., the score characterizing one group of measurements does not differ (within an acceptable margin of measurement error) from the score characterizing another group. Note the hypothesis does not state the two scores are the same; rather, it states no significant difference can be detected. Performing the statistical procedure yields a test result that helps one reach a decision that 1) the scores are not different (the hypothesis is confirmed) or 2) the difference in the scores is too great to be explained by chance (the hypothesis is rejected)" (Lehmkuhl, 1996. vol.8 Num.3).

A type 1 error is committed on rejecting nulhypothesis when in fact it is true and accepting alternative hypothesis when it is false is term as Type 11 error. "Hypothesis Testing is the art of testing if variation between two sample distributions can just be explained through random chance or not. If we must take enough precaution to see that the difference are not just through random chance. At the heart of Type 1 error is that we don't want make an unwarranted hypothesis so we exercise a lot of care by minimizing the chance of its occurrence". (STATC 141 lecture FDR)

Traditionally we try to set Type 1 error as .05 or .01 as there is only a 5 or 1 chance that variation that we are seeing is due to chance.

III. WILCOXON SIGNED RANK TEST

The wilcoxon sign rank test, is a nonparametric alternatives technique for -1 and -2 samples Student t- test in situation where the data under observation does not satisfy normality assumptions and the sizes considerably so small to be check if one exist Bellera et.al. (2010).

Although the data here need not be normal but assumes the distribution of the data sample median is symmetric in nature (Vile, J.2013).

A. THE WILCOXON SIGNED RANK TEST: ONE SAMPLE

Assuming you are to test the nullhypothesis $H_0 : \eta = \eta_0$, where η represent distribution median and η_0 is a nullhypothesised median then we computes:

$D_i = Y_i - \eta_0 \;\; \forall i$. That is to get the individual differences between all the elements in the distribution and the null hypothesis value (median) and then, $R_i = \sum_j^n I(|D_j| < |D_i| + \frac{1}{2}\sum_j^n I(|D_j| = |D_i| + \frac{1}{2}$.

Meaning having got the differences you now give them ranks in either ascending or descending order of magnitude, then $T_- = \sum_{i=1}^n I(D_i < 0)R_i$ that is sum up the ranks of the negative entities and then, $T_+ = \sum_{i=1}^n I(D_i > 0)R_i$ this mean the total ranks of the entities with positive sign (greater than or equal to 1). Other important facts:

$$\sum_{i=1}^n R_i = 1+2+\ldots+n = \frac{n(n+1)}{2} = T_- + T_+ = \frac{n(n+1)}{2}$$

this fact tells us that, the total sum of all the individual ranks of a data set of a One sample under observation (positive and negative) agrees with $\frac{n(n+1)}{2}$.

B. THE TEST STATISTIC T DEPENDS ON THE FORM OF THE ALTERNATIVE HYPOTHESIS

i. If H_1:η<η_0 (lower tailed then T=T_- (η_0).
ii. If H_1: η>η_0 (upper tailed) then T= T_+(η_0).
iii. If H_1: η≠ η_0(two tailed) then T=max{T_- (η_0),T_+ (η_0)}.

Basically for given values of n and significances levels α, critical values for T can be obtained from Statistical tables and will reject H_0 if $T_C \geq T_t$ (when T calculated is greater than T tabulated).

3.1 The Wilcoxon signed rank test: Two Sample

This is the situation where you might come across two sample population under observation i.e. If D= $X_1 - X_2$ where X_1 and X_2 have the same distribution, then it follows that the distribution of D is symmetric about zero. Thus the Wilcoxon signed rank test is often described as a test of the hypothesis that two distributions are the same that is$X_1 \sim X_2$. Let d_j denote the difference for any matched pair of observation d_j =$X_1 - X_2$ for j= 1, 2…... n. rank the absolute values of the differences, $|d_j|$ and assign any tied values the average rank. Consider the sign of d_j and let r_j = sign(d_j) and $(|d_j|)$ be the signed ranks.

The test statistics is:

i. If H_1:η<η_0 (lower tailed then T=T_- (η_0).
ii. If H_1: η>η_0 (upper tailed) then T= T_+(η_0).
iii. If H_1: η≠ η_0(two tailed) then T=max{T_- (η_0),T_+ (η_0)}.

Basically for given values of n and significances levels α, critical values for T can be obtained from Statistical tables and will reject H_0 if $T_C \geq T_t$ (when T calculated is greater than T tabulated).

IV. THE FRIEDMAN TEST

Is another type of nonparametric test which can be used to study the differences between three or more groups on circumstances where the data comes from matched pair, or repeated measures experiment, the scores of the distribution is continuous and symmetric and of course the conditions of one way ANOVA are not met.

A. IMPLEMENTATION

Assume there are n participants tested under K conditions (groups).

- ✓ First thing to do is to rank the data from lowest to highest for each group.

$$R_{ij}=\sum_{i'=1}^{k} I(Y_{i'j} < Y_{ij}) + \frac{1}{2}\sum_{i'=1}^{k} I(Y_{i'j} = Y_{ij})$$

- ✓ Sums up the individual ranks for each group:

$$R_i = \sum_{j=1}^{ni} R_{ij}$$

- ✓ Finally Friedman test statistic can be computed by the formula:

$$F_r = \frac{12}{nk(k+1)} \sum_i^k R_i^2 - 3n(k+1)$$

Where, n is number of participants, k is number of groups and R_i is sum total ranks for i group.

B. CRITICAL VALUE

If the sample size is small, critical values of the test statistic for given α, found in Friedman statistical tables. But when nk or k is large such that nk > 15 or k > 4, the probability distribution of F_r can be approximated by a χ^2-distribution with k – 1 degrees of freedom and p-value will be in form P (χ^2(k-1)≥ F_r). reject null hypothesis in favor of alternative hypothesis when the p-value is greater than or equal to F_r (calculated) statistic and then appropriate post-hoc test should be employed Dr Julie, V. (Lect. 2013. P: 76-78).

C. POST-HOC TEST

If the result obtained from the Friedman test is significance, that is if there exists difference between the conditions where the group was tested, hence you need to employ post–hoc test in order to determine where the actual differences lie. There are many post-hoc tests but the simplest one is Bonferroni test. To avoid type1 error take the significance level you are using and then divide it by the number of test you are running if, the P-value is larger than say $\frac{\alpha}{n}$ (where α is significant level and n is the number of test), then there is no difference between the number of comparison group .

V. CONCLUSION

In this paper I have discussed three of the most common nonparametric tests: the Kruskal Wallis test which is used to perform One-Way Analysis of Variance, Wilcoxon signed rank test which is an alternative to Student t-Distributions and Friedman test as an alternative to Two-Way Analysis of Variance and more importantly, conditions required of each individual test and how it can be carried out were provided and explained.

There are so many nonparametric tests. Generally speaking the nonparametric tests does not take into account the actual value of the data, rather the individual ranks position of the data in view of the above the nonparametric tests seem to have lost some important information of the data, as the ranks might not reflects the exact characteristics of the data under investigation 100 percent. Although the nonparametric tests are considered being weak but some time discovered more "robust" in cases where the data is considerably small and not normal or we are not sure of what type of data we have at hand it is better to use nonparametric tests since they can work for any population distribution.

Investigation has shown that usually the nonparametric test are only slightly less efficient than the parametric tests when the data is normal an`d they can be slightly more efficient than parametric tests when the underline normality assumptions are not met. Pitman (1948), Hodges and Lehman (1956), Chemoff and Savage (1958) they all showed that the nonparametric rank tests have desirable efficiency properties to parametric tests competitors.

REFERENCES

[1] Agresti A and Franklin C (2009).Statistics The Art and Science of Learning from Data. Pearson International Edition, 2nd Ed. Pearson Education, Inc: United States of America.

[2] Arnold B T and Emerson W J (2011) Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions. The R Journal [online] vol. 3/2, PP: 34-39: Available online at: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf [Accessed 15th February 2018].

[3] Bellera A C, Jullen M and Hanley A J (2010). Normal Approximations to the Distributions of the Wilcoxon Statistics: Accurate to what N? Graphical Insights. Journal of Statistics Education [online], Vol. 18, No.2, and PP: 1-17: Available at: http://www.amstat.org/publications/jse/v18n2/bellera.pdf [Accessed 5th February 2018]

[4] Boston University (2013). Lecture note on Test with more than Two Sample test. Available online at: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/BS704_Nonparametric7.html [accessed 17/01/2018]

[5] Dr. Vile J (MAT001, 2013). Lecture note on nonparametric tests. School of Mathematics, Cardiff University, UK.

[6] Gabiana I (2012) Freidman test, presentation. Available online at: http://www.slideshare.net/irenevictoria/friedman-test-a-presentation [accessed 9th February 2018]

[7] Mann P S (2001).Introductory Statistics, 4rd ed. John Wiley and Sons, INC: United State of America.

[8] Maurya V N, Arora D K and Maurya E A K (2013). A survey Report on Non-parametric Hypothesis Testing. International Journal of Information Technology and Operations Management [online] Vol. 1, No. 2, May 2013, PP: 29 - 40: Available online at: http://www.acascipub.com/Journals.php [Accessed 6th February 2018]

[9] Meek E G, Ceyhun O and Dunning K (2007). Comparison of the t vs. Wilcoxon Signed-Rank Test for

Likert Scale Data and Small Samples. Journal of Modern Applied Statistical Methods [online] Vol. 6, No. 1, May 2007, 91- 106: Available online at: http://digitalcommons.wayne.edu/jmasm/vol6/iss1/10 [Accessed 9th February 2018]

[10] Multiple Hypothesis Testing and False Discovery Rate STATC 141. Available online at: http://www.stat.berkeley.edu/users/hhuang/STAT141/Lecture-FDR.pdf [Acessesd 04 February 2018]

[11] Sprent P (1989). Applied Nonparametric Statistical Methods. Chapman and Hall: London.

[12] University of Sussex (2011). Lecture note on Wilcoxon Test [online]: Available online at: http://www.sussex.ac.uk/Users/grahamh/RM1web/WilcoxonHandoout2011.pdf. [Accessed 01 February, 2018]