Volume : 2 | Issue : 3 | March 2013 • ISSN No 2277 - 8160

Research Paper | Biostatistics

Solution of Nonparametric Analytic Method Kruskal Wallis Test in Health Sciences

Dr. P.V.Ubale

Associate professor in Statistics, G.S. Science, arts & Commerce College Khamgaon 444303, Distt. Buldana, Maharashtra, India

ABSTRACT

Statistics are an extremely important aspect of biology. The science of biostatistics encompasses the design of biological experiment especially in medicine and agriculture, the collection summarization and analysis of data from those experiments. In this paper we have studied the most popular nonparametric test which is a rank based test named as Kruskal Wallis test and its application is explained with the help of numerical example

KEYWORDS: Rank based test, Nonparametric test, KW test.

1] Introduction

Statistics is an important branch of biology infact they are necessary for many areas of science. It has wider applications in the field of medicine, engineering, technology bioinformatics, biostatistics economics and finance. Statistics have proven invaluable in modern biological history. In the 1930 it was evidence from the Statistical analysis that produced the modern evolutionary synthesis.

Biological studies often involve studying extremely small numbers and microscopic properties which may come from molecules and atoms and putting them into statistical form can ad understanding and simplify the process of analyzing. Biostatistics is the application of mathematical statistics to medical and biological data. Biostatistics has emerged as distinct disciplines to deal with growing applications classical statistical methodology.

The following discoveries in biology may be developed by the well known Statistician

- Genetic theory of natural selection Developed by Sir. Ronald A. Fisher who used many statistical methods to develop his work.
- Population GeneticsWere developed by Sewall G Wright by using Statistics

Pioneering work of D'arey Thompson on growth and form also help to add qualitative disciplines to biological study.The causes of evaluation and natural selection developed by J.B.S. Haldane.

Nowadays Statistics is widely used in computation, analyzing and interpretations in many fields.

The aim of present work is to give short overview of the rank based test, Kruskal Wallis test and describes the possibility of their application in biological research.

2] Nonparametric test

In this paper we are studying the most important nonparametric test which is an extension of Wilcoxon rank sum test or Mann Whitney test for comparing more than two independent groups or treatments. The Kruskal Wallis test named after William Kruskal and W. Allen Wallis (1952) is a rank based test of location equally for three or more groups may be among the most useful of available hypothesis testing procedures for behavioral and social science research.

The popularity of KW test may be attributed to its usefulness in a variety of disciplines such as education, psychology medicine and biology.

Most application of KW test is most commonly used when there is one nominal variable and one measurement variable and the measurement variable does not meet the normality assumption of an ANOVA. It is the nonparametric analogue.

If the originally data set actually consist of one nominal variable and one ranked variable one cannot do ANOVA and must use Kruskal Wallis test.

As the nonparametric generalization of the one way analysis of variance model the Kruskal Wallis test applies when the goal is to test the difference between multiple samples and the underlying population distribution are nonnormal or unknown.

Suppose that k independent samples
$x_{11}, x_{12},\ldots, x_1n_1,\ldots, x_{k1},x_{k2},\ldots, x_kn_k$ are drawn from k continuous (not necessarily normal) populations. We want to test the null hypothesis H_0 (populations are identically distributed) against Hypothesis H_1 (Population are not identically distributed).

In the KW test ranks are used instead of raw data. The Kruskal Wallis Statistic is given by

$$H = . \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1) \quad \ldots\ldots (2.1)$$

Where Rj is the sum of the ranks in the jth sample or groups, nj is the number of observations in the jth sample, k is the number of samples, N is the number of observations in the combined sample.

This Statistic is approximately distributed as statistic with k-1 degrees of freedom.

In case of tied ranks the adjustment statistic is calculated as

$$H = \qquad H = \frac{1}{S^2}\left[\sum_{j=1}^{k} \frac{R_j^2}{n_j} - \frac{N(N+1)^2}{4} \right] \quad \ldots\ldots (2.2)$$

Where

$$S^2 = \frac{1}{N-1}\left[\sum_{j=1}^{k} \sum_{i=1}^{n_j} r_{ij}^2 - \frac{N(N+1)}{4} \right] \quad \ldots\ldots (2.3)$$

Where rij is the rank of the ith observation in the jth sample.

3] Algorithm used in KW test

The procedure behind applying Kruskal Wallis test is as follows.

i) Rank the observations in the combined sample keeping the samples in ascending order of magnitude. In case of ties the identical values are given average of the ranks.
ii) Summing the ranks in each group.
iii) The Kruskal Wallis test is then compared with the statistic with k-1 d.f.
iv) In case of tied ranks the adjustment statistic is then compared with the statistic.

4] Application of Kruskal Wallis test

Let us illustrate the application of nonparametric test Kruskal Wallis test with numerical example.

The waist circumference (In cm.) measurements in a sample of young adults aged 26-32 years according to normal, impaired glucose tolerance (IGT), and type II diabetes categories are given below.

Normal:- 71.3, 85.6, 77.9, 93, 71.5

Impaired Glucose Tolerance:-
66.2, 98.4, 89, 80.5, 100.5
Type II diabetes:- 92.6, 68.3, 95.6, 97.2, 89.6

We want to test here whether the waist circumference measurements differ with respect to the three groups of young adults.

Table 4.1

Waist circumference data & their ranks					
Normal	Rank	IGT	Rank Type II	diabetes	Rank
71.3	3	66.2	1	92.6	10
85.6	7	98.4	14	68.3	2
77.9	5	89	8	95.6	12
93	11	80.5	6	97.2	13
71.5	4	100.5	15	89.6	9

The waist circumference data shown with the ranks in the table 4.1.

Research Question:- Do the normal, IGT, and diabetic young adults have the same waist circumference levels.
Setting null hypothesis
Waist circumference levels of normal, IGT, & diabetic young adults are equal.
Test Statistic

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

Computation

$$H = \frac{12}{15(15+1)}\left[\frac{30^2}{5} + \frac{44^2}{5} + \frac{46^2}{5}\right] - 3(15+1)$$

$= 0.05\ [\ 180 + 387.2 + 423.2\ ] - 48$
$= 49.52 - 48$
$= 1.52$

Table value

P value of 0.68 when compared with distribution with 2 d.f.
This indicates that there is no significant difference in the waist circumference between the group

Conclusion:-
We conclude that the waist circumference values appears to be the same between the groups.

REFERENCES

1] Kruskal W.H. (1952): A nonparametric test for the several sample problems , Annals of mathematical statistics, 23, 525-540. | | 2] Spurier J.D. (2003): On the null distribution of the Kruskal Wallis statistic, Journal of nonparametric statistics, 15(6), 685-691. | | 3] Vargha a.S. & Delaney H (1998): The Kruskal Wallis test & stochastic homogeneity, Journal of Educational & behavioural statistics , 23, 170-192. | | 4] B. Antonisamy, Solomon Christopher, P. Prasanna Samuel (2010): Biostatistics : Principles and Practice, Tata Mcgraw Hill Education Private Ltd. New Delhi |