International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321- 4767 P-ISSN: 2321-4759
www.ijmsi.org Volume 10 Issue 2 || Mar.-Apr. 2022 || PP 01-03

A Note on Kruskal Wallis Test Statistic

Anusha Ch. and Bhatracharyulu N.Ch.
Department of Statistics, University College of Science, Osmania University, Hyderabad

Abstract
Kruskal-Wallis (1952) proposed a nonparametric approach to carry out the analysis for one way classified data when the data measured on ordinal scale. In this paper an attempt is made to propose a note on test statistic.
***Keywords** or Phrases: Nonparametric method, ordinal scale, completely randomized design.*

Date of Submission: 28-04-2022 Date of Acceptance: 17-05-2022

I. INTRODUCTION

When the data is violating the basic assumptions to carry out the analysis of variance, like, normality or exact magnitudes of responses are unknown, then it is difficult for us to carry out the analysis as per the original plan of design of experiment. In such situations, we can prefer nonparametric methods for their analysis.

Several authors made attempts on developing nonparametric methods for the analysis of various experimental designs. Among them one of the popular methods used is Kruskal-Wallis test introduced by Kruskal and Wallis (1952) for the analysis of k- samples to test the equality of location parameter of those populations. In this paper a note on evaluation of test statistic by taking the ratio of variances of ranks of mean squares of treatments and errors instead of ratio of treatments to total and compared and illustrated with a suitable Design of Experiment.

II. NOTE ON PARAMETRIC & NON-PARAMERIC ANALYSIS FOR CRD

Suppose an experiment is conducted with a Completely Randomized Design layout on 'N' homogeneous experimental units with 'v' treatments, where $N= \Sigma_i n_i$ (i = 1, 2, … , v). Assume that distribution function of responses is F and is continuous with an unknown median M. Let y_{ij} (i = 1, 2, … , v, j = 1, 2, …, n_i) are the independent observed responses corresponding to i^{th} treatment applied in j^{th} replicated experimental unit.

Table-1: Classified experimental data

Treatments (T_i)	1	2	…	j	…	n_i
T_1	y_{11}	y_{12}	…	y_{1j}	…	y_{1n1}
T_2	y_{21}	y_{22}	…	y_{2j}	…	y_{2n2}
…						
T_i	y_{i1}	y_{i2}		y_{ij}		$y_{i\,ni}$
Tv	y_{v1}	y_{v2}	…	y_{vj}	…	y_{vnv}

The responses are transformed into ordinal scale by arranging them all in ascending order of their magnitude and assign ranks, 1 to N. Let R_{ij} be the rank of the response y_{ij} (i = 1, 2,…,v ; j = 1, 2, …, n_i). The rank transformed data is presented below.

Table-2: Transformed experimental data

Treatments (T_i)	1	2	…	J	…	n_i	Total	Mean
T_1	R_{11}	R_{12}	…	R_{1j}	…	R_{1n1}	$R_{1.}$	$\overline{R_{1.}}$
T_2	R_{21}	R_{22}	…	R_{2j}	…	R_{2n2}	$R_{2.}$	$\overline{R_{2.}}$

T_i	R_{i1}	R_{i2}		R_{ij}		$R_{i\,ni}$	$R_{i.}$	$\overline{R_{i.}}$

Tv	R_{v1}	R_{v2}	…	R_{vj}	…	R_{vnv}	$R_{v.}$	$\overline{R_{v.}}$

Where $\overline{R_{i.}} = \sum_{i=1}^{v} \frac{R_i}{n_i}$ the average rank of responses belongs to the i[th] treatment, and $\overline{R_{..}} = \frac{R_{..}}{N} = \sum_{i=1}^{v}\sum_{j=1}^{n_i}\frac{R_{ij}}{N} = \frac{(N+1)}{2}$ is the average rank all responses. The null hypothesis (H_0) to be tested is that there is no significant difference in the treatment effects.

H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_v = \alpha$ vs H_1: at least two of the α_i's are different

Treatment Sum of squares with (v-1) degrees of freedom is

$$\sum_{i=1}^{v}\sum_{j=1}^{n_i}(\overline{R_{i.}} - \overline{R_{..}})^2 = \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - N\,\overline{R_{..}}^2 = \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}$$

Total Sum of squares with N-1 degrees of freedom is

$$\sum_{i=1}^{v}\sum_{j=1}^{n_i}(R_{ij} - \overline{R_{..}})^2 = \sum_{i=1}^{v}\sum_{j=1}^{n_i}R_{ij}^2 - N\,\overline{R_{..}}^2 = \frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} = \frac{N(N^2-1)}{12}.$$

Then the Error Sum of Squares with (N-v) degrees of freedom is

$$\frac{N(N^2-1)}{12} - \left[\sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] = \frac{N(N+1)}{4}\left[\frac{(N-1)}{3} + (N+1)\right] - \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} = \frac{N(N+1)(2N+1)}{2} - \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i}$$

By taking the ratio of mean squares in ranks for treatments and errors, the test statistic is

The Test statistics is given by

$$Q = \frac{\frac{1}{v-1}\left[\sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right]}{\frac{1}{N-v}\left[\frac{N(N+1)(2N+1)}{2} - \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i}\right]} = \frac{(N-v)\left[4\sum_{i=1}^{v}R_{i.}^2 - n_i N(N+1)^2\right]}{2(v-1)\left[n_i(2N^3+3N^2+N) - 2\sum_{i=1}^{v}R_{i.}^2\right]}$$

The Test statistics is given by

$$Q = \frac{\sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}}{\frac{N(N+1)(2N+1)}{2} - \sum_{i=1}^{v}\frac{R_{i.}^2}{n_i}} = \frac{4\sum_{i=1}^{v}R_{i.}^2 - n_i N(N+1)^2}{2n_i(2N^3+3N^2+N) - 4\sum_{i=1}^{v}R_{i.}^2}$$

Example: The experimental data presented in Table 3 is study of the distribution of weights gains, when four types of feeds A, B, C, and D given to 20 rabbits of homogeneous group and with 5 selected for each feeding.

Table 3: Feed & Weight gains (kg)

A	3.35	3.8	3.55	3.36	3.81
B	3.79	4.1	4.11	3.95	4.25
C	4	4.5	4.51	4.75	5
D	3.57	3.82	4.09	3.96	3.82

The rank transformed data is given by

Table 4: Rank Transformed data

Treatments	Ranks					Total
A	1	6	3	2	7	**19**
B	5	14	19	10	15	**63**
C	12	16	17	18	20	**83**
D	4	8.5	13	11	8.5	**45**

(i) Then the Kruskal Wallis statistic value is

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{v}\frac{R_{i.}^2}{n_i} - 3(N+1) = \frac{12}{20\times 21}\left[\frac{19^2}{5} + \frac{63^2}{5} + \frac{83^2}{5} + \frac{45^2}{5}\right] - 3(21) = 12.68.$$

Critical value of χ^2_{3df} = 7.81. There is a significance difference the location parameters of four feedings.

(ii) The test statistic 'Q' is given by

$$Q = \frac{4\sum_{i=1}^{v}R_{i.}^2 - n_i N(N+1)^2}{2n_i(2N^3+3N^2+N) - 4\sum_{i=1}^{v}R_{i.}^2} = \frac{4(13244) - 5(20)(21)^2}{2(5)[5(16000+1200+20) - 4(13244)]} = 0.074$$

Critical value of χ^2_{3df} = 7.81, it can be concluded that, there is no significance difference.

(iii) The test statistic 'Q' is given by

$$Q = \frac{(N-v)\left[4\sum_{i=1}^{v}R_{i.}^2 - n_i N(N+1)^2\right]}{2(v-1)\left[n_i(2N^3+3N^2+N) - 2\sum_{i=1}^{v}R_{i.}^2\right]} = \frac{(20-4)[52976-44100]}{2(4-1)[5(16000+1200+20) - 2(13244)]} = 0.397$$

Critical value of χ^2_{3df} = 7.81. There is no significance difference between the feedings.

(iv) F-statistic is evaluated under the assumption of normality is F =13.77. Critical value of $F_{3,16}$ = 3.20 there is a significance difference between the feedings. It can be concluded that there is a contradiction to the Q statistic and H statistic in drawing the conclusion

References:

[1]. Hidetoshi Murakami (2006): "K-sample rank test based on modified baumgartner statistic and its power comparison", Journal of the Japanese Society of Computational Statistics, Vol. 19, pp 1-13.

[2]. Kruskal, W. H and W. A Wallis (1952): "The use of ranks in one-criterion variance analysis", Journal of the American Statistical Association. 47, 583-621.

[3]. Kruskal, W. H. (1952): "A nonparametric test for several sample problem", Annals of Mathematical Statistics, Vol. 23, pp 525-540.

[4]. Koch, G. G. and P.K. Sen (1968): "Some aspects of the statistical analysis of the mixed model", Biometrics, Vol. 24, pp 27-48.

[5]. Friedman, M. (1937): "The use of ranks to avoid the assumption of Normality implicit in the Analysis of Variance", Journal of the American Statistical Association, Vol. 32, pp 675-701.

[6]. Sen, P.K (1995a): "Nonparametric and robust methods in linear models with mixed effects", Tetra Mount.Math, J. , in press.

[7]. Sen, P.K and Puri, M.L. (1977): "Asymptotically distribution-free aligned rank order tests for composite hypotheses for general linear models", Zeit. Wahrsch. Verw, Geb. 39, pp175-186.

[8]. Sen, P.K. (1967a): "Asymptotically most powerful rank order tests for grouped data", Ann. Math. Statist, Vol. 38, pp1229--1239.

Anusha Ch. "A Note on Kruskal Wallis Test Statistic." *International Journal of Mathematics and Statistics Invention (IJMSI)*, vol. 10(02), 2022, pp. 01-03.