International Journal of Statistics and Systems
ISSN 0973-2675 Volume 17, Number 1 (2022), pp. 33-38

http://www.ripublication.com

Non-Parametric Approach for the Analysis of Completely Randomized Design

Anusha Ch. and Bhatra Charyulu N.Ch.
Department of Statistics, University College of Science, Osmania University, Hyderabad-7, India.

Abstract

Most of the techniques in Design and Analysis of Experiments are analysed based on assumptions on normality, independence, and homoscedasticity. If the assumption of normality is violated, in general, we prefer to use non-parametric approach for their analysis. Few non-parametric approach methods, like k-sample median test, Kruskal-Wallis test, Friedman test etc. are available in the literature for the analysis, when the data measured on ordinal scale.

This paper presents a non-parametric approach for the analysis of the data when the experiment is conducted with a completely randomized design layout and the observations measured on nominal scale.

Keywords: Nominal data, Sign test, Chi-square test.

1. INTRODUCTION

Suppose the experimental material is homogeneous, partition the experimental units into smallest units called ‘experimental units’. Experimental units are homogeneous i.e. there is no source of variation in the experimental units. The experimental units are allocated at random to the ‘v’ treatments for the experimentation, so that each unit has an equal chance for receiving any treatment, such a design is said to be ‘Completely Randomized Design’

The classical tests are available for testing the location parameter under the assumption of normality for single or two or more populations. If the nature of distribution is unknown, under the assumption of continuity, and the data is measured on nominal scale then, to test the location parameter, sign test is used. If the data is measured on ordinal scale for more than two populations Kruskal-Wallis test can be used.

In this paper, an attempt is made to propose to carry out the analysis for nominal data when the experiment is conducted in Completely Randomized Design layout.

2. NONPARAMETRIC APPROACH FOR ANALYSIS OF THE DATA

Suppose an experiment is conducted by allocating the 'v' treatments to the 'N' homogeneous experimental units in a completely randomised design layout. Assume, the response variable Y follows from a population with distribution F with unknown median M, where F is assumed to be continuous in the neighbourhood of M, i.e. $P[y>M] = P[y<M] = 0.5$. The hypothesis to be tested is H_0: $f_1(y) = f_2(y) = ... = f_v(y) = f(y)$, where $f_i(y)$ be the density of i^{th} population, i =1, 2, …v; and the densities are unknown.

Assume, the observed responses are measured on nominal scale and are dichotomized with ±1 and follows Bernoulli distribution with parameter 'p' under the assumption of no ties. Assume that the chance for each y takes ±1 is 0.5, i.e. $P[y > M] = P[y < M] = 0.5$.

Let y_{ij} be the binary response observed (nominal scale) on j^{th} replicate of i^{th} treatment, where i = 1, 2, …, v; and j = 1, 2, ... , n. Let y_{ij} is dichotomized as +1 and –1 like as

$$y_{ij} = +1 \quad \text{if } y_{ij} \geq M$$
$$= -1 \quad \text{if } y_{ij} \leq M.$$

The binary observed responses are classified in one-way model like as follows.

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}
…						
Tv	y_{v1}	y_{v2}	…	y_{vj}	…	y_{vnv}

Let n_i^+ and n_i^- are the number of positive and negative signs of observed responses belongs to i^{th} treatment such that $n_i^+ + n_i^- = n_i$. Let $n^+ = \Sigma_i n_i^+$ and $n^- = \Sigma_i n_i^-$ are total number of positive and negative signs of responses such that $n^+ + n^- = N$. The resulting data is presented like as follows

Table-2: +/- signs count for the experimental data

Treatment	+	-	Total
1	n_1^+	n_1^-	n_1
2	n_2^+	n_2^-	n_2
...	...	...	..
v	n_v^+	n_v^-	n_v
Total	n^+	n^-	N

Under H_0, all densities are equal and is arbitrary, i.e. the location parameter is same for the populations. The hypothesis can be expressed as:

H_0: the number of positive signs and negative signs are equally distributed.

H_1: the number of positive signs and negative signs are not equally distributed

A test statistic for the null hypothesis H_0: $p_1 = p_2 = \ldots = p_k$ unspecified where p is the parameter of Binomial can be derived from the goodness of fit test, such that exactly as in the case of k-sample median test. Under the H_0, the mean and variance of the test statistic n^+ are $E[n^+] = n/2$ and $V[n^+] = n/4$. The criterion besides sample number is now simply positive and negative signs for each sample is less than δ (small for insignificance). The test criterion is

$$Q = \Sigma_i \{ [X - n_i p_0]^2 / [n_i p_0(1-p_0)] \}$$

which is approximately the chi-square distribution with (k-1) degrees of freedom. Now the test statistics for testing equality of distributions using chi-square test is

$$Q = \Sigma_i \{ [n_i^+ - E(n_i^+)]^2 / E(n_i^+) \} + \Sigma_i \{ [n_i^- - E(n_i^-)]^2 / E(n_i^-)\}$$

Where Q follows χ^2 with (v-1) degrees of freedom.

Remarks

1. Assumptions: The observations are identically independently distributed;

distribution function F is assumed to be continuous in vicinity of median M and is unknown.

2. Theoretically no zero differences exist. But in practice zero differences can occur. The common procedure followed is simply ignore zero differences and reduce N accordingly and carryout the procedure.
3. For any $p \in H_0$, the $P[\text{sign } y_{ij} = \boldsymbol{v}) = (1/2)^N$, where $\boldsymbol{v} \in$ space of all vectors $\boldsymbol{v} = (v_1\ v_2, \ldots\ v_N)$ contains 2^N points, and $v_i = +1$ or -1. Under H_0 the distribution is symmetric.
4. The test statistic is used for $v > 2$ samples and the data are measured on nominal scale. If the number of treatments less than or equal to two, we use simply a sign test.
5. The theoretical frequencies were evaluated under multinomial, results to chi-square.
6. The most commonly used approximation of multinomial distribution is Chi squared approximation. As N tends to infinity and $p_1, p_2, \ldots\ p_k$ remain constant, the limiting form of $P(n_1, n_2, \ldots\ n_k)$ is

$$P(n_1, n_2, \ldots\ n_k) = \frac{1}{\sqrt{2\Pi N}}\left[\prod_{j=1}^{k} P_j\right]^{-1/2} exp\left[-\frac{1}{2}\sum_{j=1}^{k}\frac{(n_j - Np_j)^2}{N p_j} - \frac{1}{2}\sum_{j=1}^{k}\frac{(n_j - Np_j)^2}{N p_j} + \frac{1}{6}\sum_{j=1}^{k}\frac{(n_j - Np_j)^2}{(Np_j)^2}\right]$$

Neglecting the terms $N^{-\frac{1}{2}}$ we get

$$P(n_1, n_2, \ldots\ n_k) = \frac{1}{\sqrt{2\Pi N}}\left[\prod_{j=1}^{k} P_j\right]^{-1/2} exp\left[-\frac{1}{2}\sum_{j=1}^{k}\frac{(n_j - Np_j)^2}{N p_j}\right] = \frac{1}{\sqrt{2\Pi N}}\left[\prod_{j=1}^{k} P_j\right]^{-1/2} exp\left(-\frac{1}{2}X^2\right)$$

Where, $X^2 = \Sigma$ [Observed value – Expected value $]^2$ / Expected value

Example: Consider the following observed responses presented in Table-3, collected 34 homogeneous agricultural experimental units when four treatments are applied with a Completely Randomized Design layout.

Table 3: Experimental data

Treatment	Observations ('00 kgs)										
I	6.6	7.3	8.0	8.7	7.1	9.5	10.2	9.0	8.5	9.6	6.9
II	9.9	7.4	8.1	7.5	8.6	9.6	7.8	8.6			
III	8.4	8.8	8.9	8.6	9.9	9.1	10.1				
III	10.8	9.2	6.4	8.8	8.9	7.0	9.6	7.8			

Assume the data violates the normality condition. The median of the sample observations is 8.65 (= M). Assign a Positive sign if $y_{ij} > M$ and assign a negative sign if $y_{ij} < M$. The resulting 4×2 contingency table with observed and theoretical frequencies is presented in table 3.

Table 4: # Signs

Treatment	+1	-1	Total
I	6	5	11
II	6	2	8
III	2	5	7
IV	3	5	8
Total	17	17	34

The test statistic value Q = 3.87. critical value of $\chi^2_{0.05,\,3df} = 7.815$.

$\therefore$ It can be concluded that the value of the location parameter is same.

REFERENCES

[1] Freeman, G. H. and Halton, J. H. (1951): "Note on an exact treatment of contingency, goodness of fit and other problems of significance", Biometrika, Vol. 38, pp 141-149.

[2] Gerig, T.M. (1970): "A multivariate extension of Friedman's χ_r^2 test", Journal of the American Statistical Association, Vol. 70, pp 443-447.

[3] Gail, M. H. and Mantel, N. (1977): "Counting the Number of r x c Contingency Tables with Fixed Margins," Journal of the American Statistical Association, Vol. 72, pp 859-862.

[4] 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.

[5] Hollandar, M. G. Pledger and P.E. Lin (1974): Robustness of the Wilcoxon test to a certain dependency between samples. Annals of Statistics, Vol.. 2, pp 177-181.

[6] 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.

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

[8] 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.