Analysis of Preference Data Using Intermediate Test Statistic

***Akomolafe Abayomi .A. And**Jayeola S. Muyiwa**

***Department Of Mathematics And Statistics, Joseph Ayo Babalola University Ikeji-Arakeji,P.M.B 5006,Ilesa Osun State. Akomolafea@Yahoo.Com**
****Department Of Mathematics And Statistics, Federal Polytechnic, Ado-Ekiti.**

Abstract

Intermediate statistic is a link between Friedman test statistic and the multinomial statistic. The statistic is based on ranking in a selected number of treatments, not necessarily all alternatives. We show that this statistic is transitive to well-known test statistic being used for analysis of preference data. Specifically, it is shown that our link is equivalent to the multinomial test statistic when only one alternative is selected for ranking but to Friedman statistic when all alternatives are selected for the purpose. Having used intermediate statistic in knowing whether all the alternatives on why people choose to be educated have the same probability with respect to their preferability, it was discovered that at least one alternative is significantly preferable to the other.

Keywords:-Preference data, Friedman statistic, multinomial test statistic, intermediate test statistic.

Introduction

In nonparametric statistic, preference data consist of ranks assigned to the given alternatives by the members randomly drawn from a population. These members who may be voters, consumers or generic blocks, assign the ranks such as 1, 2, 3 to the treatment or alternatives from the most favoured one to the least favoured one. In this context, the alternatives are the choices for a question, which are given to the individuals to vote for their favourite, as a part or whole body of a questionnaire.

Let A_1, A_2, …. A_c denote *c* different alternatives. If an individual in the sample is asked to complete ranking. On the other hand, if the respondents are asked to rank a portion of the alternatives only, then we have partial or incomplete ranking. The complete ranking approach may give rise to a "psychological assessment problem". When the respondents are asked to rank all alternatives, the selection of the most favourable ones would not be a big issue in general; however, the higher order rankings f alternatives might be problematic due to the perception difficulties. The severity of the psychological assessment problem tends to increase with the number of alternatives. Therefore, if the structure of alternatives seems clearly perceptible, then the complete ranking approach may be a way to obtain preference data. Conversely, if there are too many alternatives with possible perception problems, then ranking of all these alternatives may not be easy for the respondents. In such situation, the partial ranking approach varies from 1 to the total of treatments.

Intermediate statistic is a new version of the statistic suggested by Friedman (1937) for analysis of preference data. The motivation is that a respondent may face

with a psychological difficulty when she has no assign large ranks to the options. As a panacea, intermediate statistic suggests a new method and consequently a new statistic related to preference data.

Material and methodology

A questionnaire containing nine alternatives on "why people chose to be educated" was administered on four hundred (400) students in January, 2013 from a sample drawn from the student's population of Ekiti State University Nigeria.

Specifically, these alternatives are: To read and write (A_1), to be exposed and sociable (A_2), to involve in politicking (A_3), to be relevant in the society (A_4), to get a white cola job (A_5) for effective communication (A_6), for acquisition of knowledge (A_7), for the removal of inferiority complex (A_8), and for livelihood A_9). The respondents were asked to rank their reasons preference from the most preferred to the least preferred as 1, 2 and 3. In doing this, the respondents have no difficulty in ranking their first few alternatives but possible mistakes in ranking arise due to perception problems when they are asked to rank further alternatives. Our interest is whether there is any statistically significant structure in respondent's preference regarding the reasons why people chose to be educated. Therefore, the hypothesis to be tested is that all the alternatives have the same probability with respect to their preferability.

An Intermediate Test Statistic

Assume that a random sample with the size is drawn from a given population. The respondents are asked to rank only k alternatives of A_1, A_2,…, A_c from most important to the least important. Notice that i^{th} respondent's favorite permutation is given B_1, B_{i2}, B_{i3}….. B_{ik} which is any permutation with k elements of A_1, A_2,……..., A_c based on given choice rule. The null hypothesis H_o can be described as

$$P(i^{th} \text{ respondents favorite permutation is equal to } B_{i1}, B_{i2}\ldots B_{ik}) = \frac{C}{(C-k)!} \quad \text{Where } I \in (1, 2, ., n) \qquad (1)$$

Based on the permutations given above, the ranking method can be described as follows. 1 is assigned to B_{i1}, 2 to B_{i2}, ……, k to B_{ik}; however c is assigned for the rest of all the alternatives which are not ranked or chosen by the respondents. The implication being that any non-ranked alternative will be punished more severely than those that have been ranked by the respondents. Therefore, the data in the following contingency table represent the ranked data.

Table 1

	Alternatives
Individuals	A_1, A_2……. A_c
I_1	R_{i1}, R_{i2}……. R_{1c}
I_2	R_{2i}, R_{22}……. R_{2c}
⋮	⋮
I_n	R_{ni}, R_{n2}……. R_{nc}
Total	$R_{.1}$ $R_{.2}$ ……. $R_{.C}$

In the above contingency table (I_1, I_2,., I_n) denote the respondents in the sample randomly drawn from a population and $R_{ij's}$ for I ∑ (1,2, ……..n) and j ∑ 1, 2, …, c) denotes the ranking that is based on the

explained procedure. The probability mass function is given by

$$P(R_{ij} = r_{ij}) = \begin{cases} i/c, & r_{ij} \in (1,2,\ldots\ldots,k) \\ i-k/c, & r_{ij} \in (c) \end{cases} \quad (2)$$

Where $i \leq k \leq C$, $i \in (1,2, \ldots\ldots., n)$ and $j \in (1, 2, \ldots, c)$

The formula for intermediate test statistic (s) is given as

$$s(k) = c(k) \sum_{j=1}^{c} R.j - n\left(\frac{K(K+1)}{2c} + c - k\right)2 \quad (3)$$

Where $c(k) = \dfrac{C-1}{nc\left(\dfrac{k(k+1)(2k+1)}{6c} + c(c-k) - \left[\dfrac{k(k+1)}{2c} + c-k\right]\right)^2}$

It should be noted that S (k) will have a chi-square distribution with C-1 degrees of freedom no matter what the value of k is the summary of the ranked data for k = 3.

Table 2

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
$R_{.1}$	$R_{.2}$	$R_{.3}$	$R_{.4}$	$R_{.5}$	$R_{.6}$	$R_{.7}$	$R_{.8}$	$R_{.9}$
270	**107**	**52**	**188**	**79**	**87**	**149**	**56**	**17**

Substituting 3 for k, 8 for c and the column sums in the above test statistic yields the value of the statistics, Q(k) = 1020. Based on the χ^2- approximation, a critical value for the 1% significance level is $\chi^2_{c-1,}\, r = 18.48$.

Interpretation of Result

The result strongly suggests the rejection of H_o at 1% significance level. This result is interpreted as that at least one alternative is significantly preferable to others.

Conclusion

Intermediate test statistic is a test used for the analysis of preference data. It is a test between the multinomial and Freidman test statistics. This test statistic is considered to be more general than those two. It is especially useful when the alternatives seem close to each other from respondent's perspective.

Recommendation

The test statistic used in this research work for preference data is hereby recommended as the best test statistic for the analysis of preference data.

.

.References

[1] Anderson, R.L, *Use of Contingency Tables in the Analysis of Consumer Preference Studies,* Biometrics, Vol. 15, No 4, pp. 582-590, 1959.

[2] Best, D.J., Extended Analysis for Ranked Data, Austrailian Journal of Statistics, No. 35, pp. 257-262, 1993.

[3] Dudewicz, E.J, Mishra, S.N., **Modern Mathematical Statistics**, John Wiley and Sons, New York, 1988.

[4] Friedman, M., *The Use of Planks to Avoid the Assumption of Normality Impicit in the Analysis of Variance* Journal of the American Statistical Association, No. 32, pp. 675-701, 1937.

[5] Hill, I.D., Some Aspects of Election-to-fill one seat or many, Journal of Royal Statistical Society A, No. 151, pp. 310-314.

[6] Myers, R.H., A First Course in the Theorey of Linear Statistical Models, PWS-KENT, Boston, 1991.

[7] Taplin, R.H., *The Statistical Analysis of Preference Data*, Applied Statistics, No. 4, pp. 493-521, 1997.