IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728,p-ISSN: 2319-765X, Volume 7, Issue 4 (Jul. - Aug. 2013), PP 57-61
www.iosrjournals.org

Generalized Method for Rank Determination in Rank-Ordered Statistics

[1.] Oyeka I.C.A [2]Okeh U. M
[1,]*Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State,Nigeria.*
[2]*Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki,Nigeria.*

Abstract: *This paper proposes a generalized and structured method for use for rank determination in rank-order statistics. The sampled populations may be measurements on as low as the ordinal scale and need not be continuous or numeric.*

The proposed method would readily enable the researcher assign ranks to sample observations without the need to first arrange the observations in some form as is often the case with the traditional approach in the ranking of observations. The method also provides expressions that are intrinsically and structurally formulated to enable one easily break ties and assign appropriate ranks to any tied observations if need be in a straight forward manner. The proposed method is illustrated with some sample data and shown to be often easier to use in practice than the traditional method and of more generalized and wider applicability than some other existing formulations which are often limited in their use.

Keywords*: ordinal scale, positive integers, rank order statistics, intrinsically and structurally*

I. Introduction

Rank order statistics are the basic for most non-parametric methods providing the required data as functions of the original observations for use in statistical analyses. Rank order statistics for a random sample are any set of constants which indicate the order of the observations in terms of their magnitudes or relative relationships. The actual magnitude or size of any observations is used only in the determination of its relative position in comparison with other observations in the sample data set and thereafter ignored and not used in subsequent analyses based on rank order statistics. In other words any statistical procedures based on rank-order statistics depend only on the relative magnitudes or positions of the observations in comparison with other observations in the sample.

Rank order statistics may alternatively be defined as the set of numbers which results when each original observation is replaced by the value of some any set of order-preserving function. Although theoretically any set of order preserving function may be used in the assignment of ranks to sample observations, in practice, for simplicity, only the set of the first positive integers, that is a permutation of the first set of integers is preferably used for this purpose. In assigning this set of ordered numbers or positive integers as ranks to sample observations, the traditional and usual approach has often been to first arrange the sample observations either from the smallest to the largest or largest to the smallest and then assign them the positive integers as ranks accordingly, that is either in increasing or decreasing order. This approach is however rather ad-hoc, heuristic and not systematized. Never-the-less an expression exists for the determination of the ranks that may be assigned to a set of sample observations (Gibbons, 1973,P.92).

This formulae unfortunately can only be used mostly with populations that are continuous and numeric measurements. In continuous populations no any two observations are exactly equal to each other and the probability that any two observations from such populations are exactly equal is theoretically zero. In such a situation the set of ranks obtained and assigned to sample observations using the currently existing formulae are always distinct positive integers. However in reality a set of observations do not always have different values. Some of the observations may have equal magnitudes or values and hence treated as tied observations. Strictly speaking the existing expression can not be used in breaking ties and determining the ranks of tied observations.

Several methods however exist for breaking ties between sample observations in their ranking. If the ties are few, the problem of tied observations may be resolved by dropping these tied observations and reducing sample sizes appropriately in subsequent analyses. The problem of ties if they are not too many may also be resolved by assigning tied observations their mean ranks(Freund,1992;Gibbons,1973;Hollander and Wolfe,1999,Oyeka et al,2009,Seigel,1956).Thus, if there are only few ties and they could be dropped and ignored in subsequent analysis, then the currently existing expression for the determination of ranks for sample observations if they are numeric is still applicable. If however the ties are not few and there is a need to break tied observations by assigning them their mean ranks, then the existing formulation may not be readily used for

this purpose. Furthermore, if the sampled populations are non-numeric measurements on as low as the ordinal scale, such as letter grades or scores, then the currently existing expression is of no use in assigning ranks to the observations. In this paper we propose to develop a more generalized formulation or method that would enable a researcher to systematically assign ranks to observations measured on as low as the ordinal scale whether or not the sampled populations are numeric or non-numeric. The proposed method obviates the need to require the sampled populations to be continuous or even numeric. They may be measurements on as low as the ordinal scale, and need not be continuous but may be some discrete populations.

II. The Proposed Method

Now suppose x_i is the ith observation or score in a random sample of size 'n' drawn from population X for i=1,2,…,n. Population X may be measurements on as low as the ordinal scale and need not be continuous or numeric. To develop a more generalized method for use in determining and assigning ranks to sample observations measured on as low as the ordinal scale, we may let,

$$u_{ij} = \begin{cases} 1, \textit{if the jth observation or score } x_j \textit{ is} \\ \textit{at least as high (good, large, great, serious) as the ith observation} \\ \textit{or score } x_i; x_j \geq x_i \\ 0, \textit{if the jth observation or scoring } x_j \textit{ is lower(worse, smaller,} \\ \textit{less, less serious) than the ith observation or score } x_i; x_j < x_i \end{cases} \quad 1$$

$$\textit{for } i, j = 1, 2, \ldots n$$

Note that u_{ij} of Equation 1 is defined and applicable to all data sets whether or not continuous and whether or not data numeric provided they are measurements on at least the ordinal scale.

Then the rank order statistic $r(x_j) = r_j$ that is the rank assigned to the jth observation or score in the ranking of the 'n' sample observations from the largest or highest to the smallest or least may be determined from the equation.

$$r(x_j) = r_j = \sum_{i=1}^{n} u_{ij} \quad 2$$

$$\textit{for some } j = 1, 2, \ldots, n$$

Note that unlike the approach with the traditional or usual ad-hoc method for determining the ranks for sample observations, the approach used in Equations 1 and 2 do not require the researcher to first arrange the observations from the lowest to the highest or highest to the lowest before assigning them ranks. The observations are rather ranked as presented without any problems. Research interest may also be in resolving the problem of ties between sample observations by assigning mean ranks to observations tied in values. Now suppose x_j assigned the rank $r(x_j) = r_j$ using Equation 2 is tied in value with some t-1 other observations or scores in the sample ,then the mean or average rank of 't tied observations with rank 'j' here designated notationally as av.r (r_j, t) may be determined from the Equation

$$\textit{average } (r_j, t) = r_j - \frac{(t-1)}{2} \quad 3$$

$$\textit{for } j, t = 2, 3, \ldots, n$$

In other words in general note that the n sample observations may not all be distinct. They may consist of some k groups of different observations with each group containing observations that have equal values or tied observations. Thus in this situation the rank r_j of the jth observation would then in effect be the rank r_h common to some t_h ,say, observations tied in value in the hth group of tied observation for h=1,2,…,k. This means that the jth observation x_j which now also may be taken as the hth observation in the hth group of tied observations is tied in score with $t_h - 1$ other observations in that group. Hence the mean or average rank $avr(r_h; t_h)$ of the hth observation tied in value with some $t_h - 1$ other observations in the hth group of tied observations, is calculated as

$$avr(r_h; t_h) = r_h - \frac{(t-1)}{2} \quad 4$$

Note that if $t_h = 1$,that is if the hth observation with rank r_h is not tied in value with any other observations in the hth group of tied observations in the sample, then $r_j = r_h, for\ j = 1,2,...,n$ and some h=1,2,…,k.

It is noteworthy that unlike the traditional or usual approach to the breaking of ties between tied observations by assigning them their mean ranks, no particular and deliberate effort is made here in assigning mean ranks using the proposed method. Equation 4 has been intrinsically and structurally formulated to achieve this need with respect to tied observations or scores.

Now note that since the ranks or mean ranks $avr(r_h;t_h)\ forh = 1,2,..,h$ the ranks assigned to the 'n' sample observations are each generally a member of the set of the first 'n' positive integers, their weighted sum must be the sum of there 'n' positive integers namely $\frac{n(n+1)}{2}$ and their mean $\bar{r}$ must also be $\frac{n+1}{2}$,the average or mean value of these 'n' positive integers. In other words specifically the weighted sum of average ranks $avr(r_h;t_h)$ for all sample observations where the weights are the values of ' t_h 'the number of tied observations associated with their hth group of tied observations in the sample is equal to the total sum of the ranks namely $\frac{n(n+1)}{2}$ assigned to the 'n' sample observations. Expressed notationally we have that the sum of the ranks 'r' assigned to 'n' sample observations namely $S(r;n) = \sum_{h=1}^{k} t_h avr(r_h;t_h)$ has the same value as $\frac{n(n+1)}{2}$,where t_h is the number of observations in the hth group or set of tied observations and $avr(r_h;t_h)$ is the average of the ranks r_h calculated for the group of observations x_h which is tied in value with $t_h - 1$ other observations in the sample for h=1,2,…,k groups. Now from Equation 4 we have that

$$S(r;n) = \sum_{h=1}^{k} t_h avr(r_h;t_h) = \sum_{h=1}^{k} t_h \left(r_h - \frac{1}{2}(t_h - 1) \right) or$$

$$S(r;n) = \sum_{h=1}^{k} t_h . r_h - \frac{1}{2}\sum_{h=1}^{k} (t_h^2 - t_h) \qquad 5$$

The corresponding mean of the ranks is

$$S(\bar{r};n) = \frac{S(r;n)}{\sum_{h=1}^{k} t_h} = \bar{r} = \frac{n+1}{2}$$

the mean or average of ranks assigned to the 'n' sample observations which from Equation 5 yields

$$S(\bar{r};n) = \frac{\sum_{h=1}^{k} t_h . r_h - \frac{1}{2}\sum_{h=1}^{k} (t_h^2 - t_h)}{\sum_{h=1}^{k} t_h} \qquad 6$$

Where $\sum_{h=1}^{k} t_h = n,$ the total sample size. Equations 5 and 6 provides alternative expressions for the calculation of the sum and mean respectively of the ranks assigned to 'n' sample observations or scores directly from first principles instead of relying on mere memory recall.

III. Illustrative Example

We here illustrate the proposed method with measurements on the ordinal scale namely letter grades. The letter grades by a random sample of 12 medical students who took an introductory course in medical statistics in a certain University are C,B,F,A,B,E,D,C,B,D,B,C. We use these letter grades to illustrate the proposed method and the assignment of ranks to non-numeric measurements on the ordinal scale as well as the resolution of ties between observations by assigning tied observations their mean ranks. To determine the ranks

to be assigned to letter grades earned by students we may first apply Equations 1 to the above data to obtain the 1s and 0s namely, values of r_h which for simplicity are here presented in tabular form (Table 1)

Table 1: Values of u_{ij} (Equation 1) for student letter grades

		$x_{j\,1}$	2	3	4	5	6	7	8	9	10	11	12	$(r(x_i))$
x_i		C	B	F	A	B	E	D	C	B	D	B	C	
1	C	1	1	0	1	1	0	0	1	1	0	1	1	8
2	B	0	1	0	1	1	0	0	0	1	0	1	0	5
3	F	1	1	1	1	1	1	1	1	1	1	1	1	12
4	A	0	0	0	1	0	0	0	0	0	0	0	0	1
5	B	0	1	0	1	1	0	0	0	1	0	1	0	5
6	E	1	1	0	1	1	1	1	1	1	1	1	1	11
7	D	1	1	0	1	1	0	1	1	1	1	1	1	10
8	C	1	1	0	1	1	0	0	1	1	0	1	1	8
9	B	0	1	0	1	1	0	0	0	1	0	1	0	5
10	D	1	1	0	1	1	0	1	1	1	1	1	1	10
11	B	0	1	0	1	1	0	0	0	1	0	1	0	5
12	C	1	1	1	1	1	0	0	1	1	0	1	1	8
$(r(x_j))$		7	11	1	12	11	2	4	7	11	4	11	7	

The total column ranks of Table 1 $r(x_j) = r_j$ read column wise are the ranks of the letter grades, ranked from the highest grade A assigned the rank 12 to the lowest grade F assigned the rank 1;while the total row ranks $r(x_i) = r_i$ read row-wise are the ranks of the same letter grades if ranked from the highest grade A assigned the rank 1 to the lowest grade F assigned the rank 12.Now using the results of Table 1 in Equations 2 and 3 or 4 we obtain the assigned ranks and the mean or average ranks of the letter grades. The results are presented in Table 2.

Table 2.Assigned ranks for students letter grades of Table 1.

Letter grades(h)	Rank $(r(x_h) = r_h)$	No of ties in grade (t_h)	Avr (r_h, t_h) $\left(r_h - (\frac{t_h - 1}{2})\right)$	Traditional ranking procedure $(r(x_h))$	Eighted ranks $(t_h.avr(r_h : t_h))$
C	7	3	6	6	18
B	11	4	9.5	9.5	38
F	1	1	1	1	1
A	12	1	12	12	12
E	2	1	2	2	2
D	4	2	3.5	3.5	7
Total					

Note as shown above that if the grades and in general sample observations are rather ranked from the least or lowest grade F to the highest or best grade A then the row totals $r(x_i)$ of Table 1 may be used to determine the ranks and the mean or average ranks of these grades in their ranking. No matter how the ranks and mean or average ranks are obtained we would always have using the results of Table 2 in equation 5 as shown in the last column of Table 2

$S(r;n) = 3(6) + 4(9.5) + 1(1) = 1912) + 1(2) + 2(3.5) = 78 = \frac{12(12+1)}{2} = \frac{n(n+1)}{2}$, the sum of the set of the first 'n' positive integers used in the ranking of the n=12 observations. Similarly from Equation 6 we have that the mean of the ranks is $S(\bar{r};n) = \frac{78}{12} = 6.5 = \bar{r} = \frac{12+1}{2} = \frac{n+1}{2}$, the mean of the ranks, that is the mean of the set of the first 'n' positive integers used in the ranking of the n=12 observations. Finally note also from Table 2 that the values of the ranks obtained for the observations using the present more generalized procedure yields the same values of ranks that would have been obtained using the traditional approach to the determine of ranks of sample observations. However the method presented in Gibbons (1973;p.92) cannot be used to obtain the same results because as already pointed out above the method is appropriate for use only with

continuous populations that are numeric measurements. The present approach is thus more generalized better structured formulated and often simpler and quicker to use in practical applications than the traditional method and the currently existing expressions for rank determinations in rank-order statistics.

IV. Summary And Conclusion

We have in this paper proposed, developed and presented a method for the determination and assignment of ranks to sample data. The method is intrinsically and structurally formulated in such a way as would enable one break ties and assign tied observations their mean ranks without the need to resort to the often lengthy traditional approach. The method is of much more generalized applicability in that unlike some other existing methods it can be used with any set of data measured on at least the ordinal scale whether or not the populations of interest are continuous and whether or not they are numeric. The proposed method is illustrated with some sample data and shown to be generally more useful in a variety of situations than some other existing methods.

Reference

[1]. Gibbons, J. D.: Non- Parametric Statistical. An Introduction; Newbury Park: Sage Publication 1993
[2]. Freund, R.J.1972.Some observations on regressions with grouped data.Amer.Statist.25(3):29-30.
[3]. Hollander, M. and Wolfe, D.A.(1999): Non-Parametric Statistical Methods (2nd Edition). Wiley Interscience, New York Freund, R.J.1971.Some observations on regressions with grouped data.Amer.Statist.25(3):29-30.
[4]. S. Siegel, "Non-Parametric Statistics for the Behavioral Sciences," McGraw-Hill Series in Psychology, New York,1956
[5]. C. A. Oyeka, C. E. Utazi,C.R.Nwosu,P.A.Ikpegbu G. U. Ebuh, H. O. Ilouno and C. C. Nwankwo, "Method of Analysing Paired Data Intrinsically Adjusted for Ties," *Global Journal of Mathematics*, Vol. 1, No. 1, 2009, pp. 1-6.