

Simplified Method of Rank Determination In Rank Order Statistics

OYEKA, C.A; *OSUJI, G.A and NWANKWO, C.C.
DEPARTMENT OF STATISTICS, NNAMDI AZIKIWE UNIVERSITY, AWKA, NIGERIA.

*Address for Correspondence OSUJI, G.A. E-mail: george.osuji99@yahoo.com

ABSTRACT

This paper proposes a simplified and formatted method for systematically assigning ranks within and between sample observations drawn from several populations: These populations may be measurements on as low as the ordinal scale and need not be continuous or even numeric. The method is formulated to intrinsically and structurally directly break any possibilities between observations both within and between sample observations and directly assign observations tied in one value their mean ranks without any additional calculations. The proposed method is illustrated with some sample data and shown to be relatively easier to use in practical applications than other existing methods that can possibly be used for the same purpose.

KEYWORDS: Rank-order, Ties, observations, Ordinal Scale, Intrinsically Adjusted

INTRODUCTION

Sometimes available sample data may be measurements on as low as the ordinal scale that are often not appropriate for analysis using parametric methods or the researcher may in any case prefer to use more robust non parametric methods, requiring the conversion of the sample data into ranks and using these ranks in subsequent analyses. This means in practical applications that the researcher would need to convert the available sample data into ranks that reflect the relation or position of the observations relative to one another in terms of their magnitudes using a set of ordered positive intergers.

In assigning this set of ordered positive intergers or numbers as ranks to sample observations, the tradition 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 intergers as ranks accordingly, that is either in increasing or decreasing order. This approach is however rather ad-hoc, heuristic, not systematized and formatted, and most often time-consuming. Fortunately a formatted expression exists for the determination of the ranks that may be systematically assigned to a set of sample observations that are continuous measurements (Gibbons 1973, pg 92). These formulae unfortunately as noted above however can only be used with populations that are continuous and numeric measurements. It can not be used with measurements on the ordinal scale or with populations with tied observation since the method can not possibly be used to break ties and assign ranks of same form to tied observations.

Several methods never-the-less exist for breaking ties, if they occur between and among sample observations in their ranking. These include dropping tied observations if they are few and reducing sample sizes accordingly; assigning tied observations randomly to one of the two classes or groups into which the sample observations may have been dichotomized by some index; or assigning observations tied in one value their mean ranks (Siegel 1956; Gibbons 1973; Freund 1992; Hollander and Wolfe 1999; Oyeka etal 2009; Oyeka and Uche 2013)

Hence if there are only few ties and they could be dropped and ignored in subsequent analysis, then the current existing expression for the determination of ranks for sample observations if they are numeric is still applicable. However Oyeka (2013) has developed a generalized method for breaking ties between and among sampled populations that does not require the dropping of these ties and may be used with populations on as low as the ordinal scale of measurement. But the approach is rather cumbersome and often time consuming in practical applications because of the need to determine the frequencies of tied observation which are then used in determining the ranks to be assigned to these observations.

In this paper, we propose and present a simplified method for the determination of ranks to be assigned to sample observations drawn from several populations which is also intrinsically and structurally formulated to enable the breaking of ties between and among sampled populations by assigning mean rank directly to each set of observations tied in one value.

THE PROPOSED METHOD

Suppose X_i , is the i^{th} observation or score in a random sample of size n_i drawn from population X_i for $i = 1, 2, \dots, n_i; i = 1, 2, \dots, c$; where ' c ' is the number of sampled populations which may be measurements on as low as the ordinal scale, and need not be continuous or even numeric. Research interest is to develop a statistical method or procedure to systematically assign ranks to observations in each of the ' c ' samples when all

the $n = \sum_{i=1}^c n_i$ sample observations are pooled and assigned ranks for possible use in further statistical analysis.

To do this we may pool the 'c' samples together in any desired way or as presented. Now let x_k be the k^{th} observation in the pooled sample of 'n' observations, for $k = 1, 2, \dots, n$

Let

$$ukj = \begin{cases} 1 & \text{if } k = j \text{ or for } k \neq j, X_k \text{ is higher (better, larger, more serious)} \\ & \text{than } X_j; \text{ or } X_k > X_j \\ \frac{1}{2} & \text{if } X_k \text{ is the same score (equal to as same as) } X_j, \text{ or} \\ & X_k = X_j \\ 0 & \text{if } X_k \text{ is lower (worse, smaller less serious) than } X_j; \\ & \text{or } X_k < X_j; \end{cases} \quad 1$$

for $k, j = 1, 2, \dots, n$.

Note that ukj of Eqn 1 is defined and applicable to all data sets irrespective of whether or not they are continuous or numeric provided they are measurements on at least the ordinal scale. Note also that Eqn 1 does not require that the sample observations or scores being analyzed be presented or arranged in any predetermined or prespecified form or order, before assignment of ranks.

Now let:

$$\prod_j^+ = P(ukj = 1); \prod_j^c = P(ukj = \frac{1}{2}); \prod_j^0 = P(ukj = 0) \quad 2$$

Where

$$\prod_j^+ + \prod_j^c + \prod_j^0 = 1 \quad 3$$

Now using Eqn 1 we have that the rank order statistic $r(x_j) = r_j$, that is the rank assigned to the j^{th} observation or score in the ranking of the 'n' sample observations from the largest or highest assigned the highest rank to the smallest or least assigned the lowest rank may be determined from the expression

$$W_j = r(x_j) = r_j = \sum_{k=1}^n ukj \quad 4$$

For some $j = 1, 2, \dots, n$.

Equation 4 enables the researcher assign unique ranks to the 'n' sample observations or scores if these observations are all distinct. However the 'n' sample observations may not always all be distinct. They may consist of some 'g' groups of different observations with each group containing observations that have equal values and hence treated as tied observations. To break ties within each of these groups of tied observation and assign observations in each group their mean ranks, we have from Eqns 1 and 2 that the expected value and variance of ukj are respectively.

$$E(ukj) = \prod_j^+ + \frac{1}{2} \cdot \prod_j^c \quad \text{Var}(ukj) = \prod_j^+ + \frac{1}{4} \prod_j^c - \left(\prod_j^+ + \frac{1}{2} \prod_j^c \right)^2 \quad 5$$

Also from Eqns 2 and 4 we have that the expected value of $W_j = r(x_j) = r_j$ is

$$E(W_j) = E(r(x_j)) = \sum_{j=1}^n E(ukj)$$

OR

$$E(W_j) = E(r(x_j)) = n \left(\prod_j^+ + \frac{1}{2} \cdot \prod_j^c \right) \quad 6$$

Now \prod_j^+ , \prod_j^e and \prod_j^0 are respectively the probabilities that the observation or value of interest is that of the j^{th} randomly selected subject himself or if not, the probability that j^{th} subject score is higher (better, greater, more serious), the same as (equal to, as serious as) or lower (worse, smaller, less serious) than the scores by all other subjects in the combined sample for $j=1,2,\dots,n$.

Their sample estimates are respectively

$$\prod_j^+ = \frac{f_j^+}{n}; \quad \prod_j^e = \frac{f_j^e}{n}; \quad \prod_j^0 = \frac{f_j^o}{n} \tag{7}$$

Where f_j^+ , f_j^e and f_j^o are respectively the number of 1s, $\frac{1}{2}$ s, and 0s in the frequency distribution of the 'n' values of these numbers on uk_j , for either k or $j=1,2,\dots,n$. In other words f_j^+ , f_j^e and f_j^o are respectively the number of times the observation or value of interest is that of the j^{th} randomly selected subject himself or if not, the number of times that j^{th} subject score is higher (better, greater, more serious), the same as (equal to, as serious as), or lower (worse, smaller, less serious) than the scores by all other subjects in the combined sample for $j=1,2,\dots,n$. Hence from Eqns 6 and 7 we have that the sample estimate of the rank assigned to x_j , the observations or score by the j^{th} subject in the ranking of the 'n' observations in the combined or pooled sample observations is

$$W_j = r(x_j) = n \left(\prod_j^+ + \frac{1}{2} \prod_j^e \right) = f_j^+ + \frac{1}{2} f_j^e \tag{8}$$

Eqn 8 would readily enable the research determine the rank or mean rank to be assigned to each observations in each of the 'C' sample Separately when these samples are combined and ranked together as one common sample both overall and for each of the c samples.

Thus the present method enables one easily and quickly systematically assigns ranks as well as breaks ties and assigns mean ranks to tied observations both within and between samples drawn from several populations for possible use in further statistical analysis. Although both methods yield the same results, as already noted above the traditional method of rank determination is relatively more difficult and time consuming to use in practical applications and may hence be less cost effective.

Using Eqn 1 enables one systematically assign ranks to sample observations from the highest (best, largest, most preferred) assigned the highest rank to the lowest (worst, smallest, least preferred) assigned the lowest rank in the preferential ranking of the observations if uk_j of Eqn 1 is summed over k , that is row-wise for each 'j'. Similarly the observations are ranked from the lowest (worst, smallest, least preferred) assigned the lowest rank, to the highest (best, largest, most preferred) assigned the highest rank if uk_j is summed across 'j' column-wise, for each k .

An additional advantage of the present method over other existing methods for rank determination in rank order statistic is that it simultaneously provide estimates of the probabilities that a randomly selected subject performs higher (better) as well as or lower (worse) than all other subjects in the sampled population(s) which could be most helpful in program planning and implementation.

Illustrative Example

An introductory course in Bio-statistics has three sections taught by the three different instructors. A researcher interested in determining whether students perform equally well in the course under the three instructors collected letter grades awarded by the course instructors to random samples of 9, 13 and 11 students who took the course in the three sections as presented in table 1.

Table 1: letter grades earned by students under three instructors

	Scores												
Section 1:	E	A ⁺	A	C	A ⁻	E	A ⁻	F	C ⁻				
Section 2:	B ⁺	B	C ⁻	A ⁻	A ⁺	D	A	D	A ⁻	A	B ⁺	C	C ⁺
Section 3:	F	D	C ⁻	C	B ⁻	A ⁻	A ⁻	B ⁺	C	D	D		

	Ranks (traditional method)												
Section 1:	30.5	1.5	4	19.5	8.5	30.5	8.5	32.5	23				
Section 2:	13	15	23	8.5	1.5	27	4	27	8.5	4	13	19.5	17
Section 3:	32.5	27	23	19.5	16	8.5	8.5	13	19.5	27	27		

To use the proposed method to determine the ranks to be assigned to the sample observations, or letter grades of table 1 we would first pool these grades into one combined sample of size 'n' =33 in any form or order.

But for simplicity we will here list and present the grades by course instructor for the purpose of their ranking.

Table 3 presents summary values of the ranks assigned to the letter grades using the results of table 2.

Table 3: Ranks, W_j (Equation (8)) of letter grades from table 2.

Letter grade (j)	f_j^+	f_j^e	$W_{j=r_j=f_j^+ + 1/2 f_j^e}$
A ⁺	1	1	1.5
A	3	2	4
A ⁻	6	5	8.5
B ⁺	12	2	13
B	15	0	15
B ⁻	16	0	16
C ⁺	17	0	17
C	18	3	19.5
C ⁻	22	2	23
D	25	4	27
E	30	1	30.5
F	32	1	32.5

The ranks that would be assigned to the sample observations, that is letter grades of table 1 based on the ranks of these grades of table 3 are presented in table 4.

Table 4: Ranks of the sample data of table 1 based on Eqn 1

Section 1		Section 2		Section 3	
Grade	rank	Grade	rank	Grade	rank
E	30.5	B ⁺	13	F	32.5
A ⁺	1.5	B	15	D	27
A	4	C ⁻	23	C ⁻	23
C	19.5	A ⁻	8.5	C	19.5
A ⁻	8.5	A ⁺	1.5	B ⁻	16
E	30.5	D	27	A ⁻	8.5
A ⁻	8.5	A	4	A ⁻	8.5
F	32.5	D	27	B ⁺	13
C ⁻	23	A ⁻	8.5	C	19.5
		A	4	D	27
		B ⁺	13	D	27
		C	19.5		
		C ⁺	17		

Notice from table 4 that the ranks assigned to the sample scores in each section using the proposed rank determination method are the same as would be obtained using the traditional method of ranking as shown in table 1 where the observations are ranked from the first grade A⁺ tied at rank 1.5 to the worst grade F tied at rank 32.5.

However, the proposed method is clearly more instructive, systematic and statistical. An alternative method that would produce the same results in practical applications using the proposed method if one does not wish to tabulate the values of u_{kj} of Eqn 1 in a spread-sheet would be to list the sample observations again in any desired form. Then if one is interested in ranking the sample observations from say, the highest (best, first, largest, most performing) assigned the highest rank to the lowest (worst, last, smallest, least performing) assigned the lowest rank, one would after pooling the observations together into one combined sample determine for each observation how many other observations in the combined sample are higher (better, larger) than it and also determine how many other observations in the combined sample are tied in value with that observation. Then the rank to be assigned to this observation in the combined ranking is the sum of the number of observations higher than it and is plus one half of the number of other observations tied in value with it. This alternative approach

clearly yields the same rank for j^{th} observation in the combined sample as $W_j = r(x_j) = f_j^+ + 1/2 f_j^e$ of

Equation 8.

SUMMARY AND CONCLUSION

This paper has presented and discussed a simplified method to use in assigning ranks to observations within and between samples drawn from several populations that need not be continuous or numeric but may be

measurements on as low as the ordinal scale. The method is intrinsically and structurally formulated to directly and systematically break any possible ties within and between sample observations and assign any set of tied observations their mean ranks. The proposed method is illustrated with some sample data and shown to be much easier to use in practical applications than other existing methods even when such methods can possibly also be used for the same purpose.

REFERENCES

1. Freund, J.E. (1992): Mathematical statistics (5th edition), Prentice-Hall International Editions, USA
2. Gibbon, J.D. (1973) non parametric statistical inference "McGraw-Hills book Company, New York.
3. Hollander, M. and Woife, D.A. (1999); Non parametric statistical methods (2nd edition) Wiley-Inter science, New York.
4. Oyeka, I.C.A and Okeh, U.M. (2013) Estimation of subject specific index of relative performance in "K" samples. American Journal of theoretical and applied statistics vol. 2 pg 154-165
5. Oyeka, C.A. etal (2009). A method of analyzing paired data intrinsically adjusted for tie Global Journal of maths and statistics vol. 1, pg.1-6
6. Oyeka, C.A. (1996): An Introduction to applied statistical methods. Nobern avocation publishing company, Enugu-Nigeria.
7. Siesel, S. (1956): Non-parametric statistics for behavioral sciences," McGraw-Hill book company, New York.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** <http://www.iiste.org/journals/> The IISTE editorial team promises to review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Recent conferences: <http://www.iiste.org/conference/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

