Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

SIMPLIFIED AND ALTERNATIVE APPROACH TO MODIFIED RIDIT ANALYSIS

[1]Matthew Chukwuma Michael; [2]Oyeka, I. C. A. and [3]Eriobu Nkiru Obioma; [4]Ehiwario, Jacob Chinedum
[1,4]Department of Mathematics and Statistics, Faculty of Science, University of Delta, Agbor State, Delta State, Nigeria
[2,3]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azkiwe University, Awka, Anambra State, Nigeria.
Corresponding Author: megawaves4life@yahoo.com

Abstract: This paper proposes a simplified and alternative modified method of Ridit Analysis for qualitatively ordered groups of sample data. The proposed method enables the estimation of the probabilities that a randomly selected subject from one sampled population is at a more serious (higher, worse), the same (equal) or less serious (lower, better) condition or state than a randomly selected subject from another sampled population. Test statistics were developed which enables the testing not only of the null hypothesis, if desired, that subjects at a given condition of seriousness in one population is at a more serious (higher, worse) condition than subjects at all levels or conditions of seriousness in the other population, but also the null hypothesis that subjects from one population are on the average in a more serious (higher, worse) condition than subjects from the other population. The method was illustrated with some sample data and the proposed test statistic was shown to be relatively more efficient and hence likely to be more powerful than the corresponding earlier modified test statistic.

KEYWORDS: Simplified, Modified, Alternative, Ridit Analysis, natural ordering and, qualitative variable.

INTRODUCTION

Suppose that sample data were drawn from a number of populations each of which is assumed to have in-built qualitatively ordered categories or classes. For example, suppose we have random samples of patients of a certain disease or accident victims by age and sex whose conditions are ordered from dead to critical, severe, moderate, improved, most improved, well, etc. Although these gradations may be coarse, discrete, and still finite, they are nevertheless more descriptive and exhaustive than merely using some dichotomous classifications such as none or all, yes or no, present or absent, etc. which are fairly crude and not fully descriptive. To compare these samples and reach clear conclusions is often difficult. However, in the above and similar cases, the grading of the degree of seriousness is subjective and may not be reliable. Furthermore, it is difficult to find a readily interpretable summary index for such data set and to make comparisons among different samples in an intelligent way. The conventional chi-square analysis may be performed, but important information on the natural ordering of the categories would be lost.

A frequently employed procedure is to number the categories from say, 0, for the least serious to some highest number for the most serious, calculate means and standard deviations and then apply the usual t −test or analysis of variance. There is, however, also a problem with this approach. The assignment of ordered numerical codes with equal spacing to the various categories of the variable under study is often arbitrary. It is a device that defines a metric on the categories of a qualitative variable which may or may not represent the true pattern of relationships among these categories.

A probably preferable approach is to fit an ordinal dummy variable multiple regression model to the data and use the usual regression technique in the analysis. This approach does not require the assumption of any form of distribution for the data and does not attempt to

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

quantify the categories but rather works with their natural ordering. Bross has also developed a similar technique called ridit analysis where ridit is an acronym for "relative to an identified distribution" of the proportions or frequencies over the various ordered categories of some chosen standard reference population [1]. Relative in the sense that proportions or frequencies of occurrence of observation in the various ordered categories of a population of interest, are compared with the proportions or frequencies in the corresponding ordered categories of the reference population or group. In ridit analysis, virtually, the only assumption made is that the discrete categories represent intervals of an underlying but unobservable continuous distribution but no assumption is made about normality or any other form for the distribution [2]; [3].

However, as [3] and [4] aptly observed, a problem with bross ridit analysis is the implicit assumption that the reference group is a population. In Bross' ridit analysis, interchanging the roles of the reference and standard groups often affects the results of the analysis, especially when the sizes of the samples drawn from these two groups are quite disparate since the standard deviations of the resulting test statistic may differ widely. In their own approaches, [3] and [4] developed modified methods of ridit analysis in which any of the sampled groups could serve as either the reference or standard group and none of these groups need to be a population but could be random samples drawn from some larger populations. Although these modified methods and the resulting test statistics are relatively more efficient and powerful than the Bross' original test statistic, they nevertheless have the problem of unacceptably high variance, a problem that would still lead to some loss of power and often misleading conclusions.

This paper therefore presents a simplified and alternative approach to the modified ridit analysis developed by [3] and [4] that is relatively more efficient with smaller variance and hence more powerful than the earlier modified methods.

THE PROPOSED METHOD

Suppose $C_1, C_2, \cdots, C_k$ are the k graded levels of the seriousness of an event or situation, injury, health status, performance of the economy, general level of socioeconomic development or condition, level of corruption, pollution and degradation of the environment, deterioration of standard of living or socio and cultural norms, etc. into which subjects, objects, items or entities drawn from two or more populations have been grouped or classified. It is here assumed for simplicity but without loss of generality that these k groups or classifications have been ordered or arranged from the least serious C_1 to the most serious C_k such that if C_l succeeds C_j, then j preceeds l and C_l is considered a more serious level than C_j, for $j, l = 1, 2, \cdots, k$.

In particular, in the case of two populations X and Y, suppose f_{lx} and f_{hy} are respectively the number of subjects or items from populations X and Y at C_l and C_h levels of seriousness of condition, these observations are not initially in their summarized form but presented as raw scores for analysis that I_{ilx} and I_{jhy} are respectively the scores or information that indicate that the i^{th} randomly drawn subject from population X is in condition l and that the j^{th} randomly drawn subject from population Y is in level h of condition of seriousness, for $i = 1, 2, \cdots, n_x$; $j = 1, 2, \cdots, n_y$; and $i, h = 1, 2, \cdots, k$.

The total number of sampled subjects or observations from populations X and Y are respectively $n_x = \sum_{l=1}^{k} f_{lx}$ and $n_y = \sum_{h=1}^{k} f_{hy}$.

The sampled data from two populations may be presented in summary form as in table 1.

To develop the proposed simplified method, let again assume but without loss of generality that populations Y and X are respectively the reference and comparison populations,

$$U_{ij} = \{1, if\, I_{ilx}\ is\ higher\ than\ I_{jhy}\ 0, if\ I_{ilx}\ is\ equal\ to\ I_{jhy}\ -1, \quad if\ I_{ilx}\ is\ equal\ to\ I_{jhy} \qquad (1)$$

For $i = 1, 2, \cdots, n_x; j = 1, 2, \cdots, n_y; l, h = 1, 2, \cdots, k$

Let

$$\pi_x^+ = P(U_{ij} = 1);\ \pi_x^0 = P(U_{ij} = 0);\ \pi_x^- = P(U_{ij} = -1) \quad (2)$$

where

$$\pi_x^+ + \pi_x^0 + \pi_x^- = 1 \quad (3)$$

Define

$$W_x = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} U_{ij} \quad (4)$$

Now

$$E(U_{ij}) = \pi_x^+ - \pi_x^-; Var(U_{ij}) = \pi_x^+ + \pi_x^- - (\pi_x^+ - \pi_x^-)^2 \quad (5)$$

From Equation 4 we have that the expected value of W_x is

$$E(W_x) = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} E(U_{ij}) = n_x n_y (\pi_x^+ - \pi_x^-) \quad (6)$$

The corresponding variance is

$$Var(W_x) = n_x n_y (\pi_x^+ + \pi_y^- - (\pi_x^+ - \pi_x^-)^2) \quad (7)$$

Now, π_x^+, π_x^0 and π_x^- are respectively the probabilities that a randomly selected subject from population X is on the average in a more serious (worse or higher), the same (equal) or less serious (better or lower) condition than a randomly selected subject from population Y. The sample estimate of these probabilities are respectively

$$\hat{\pi}_x^+ = \frac{f_x^+}{n_x n_y};\ \hat{\pi}_x^0 = \frac{f_x^0}{n_x n_y};\ \hat{\pi}_x^- = \frac{f_x^-}{n_x n_y} \quad (8)$$

where f_x^+, f_x^0 and f_x^- are respectively the total number of 1s, 0s and -1s is the frequency distribution of the $n_x n_y$ values of these numbers in $U_{ij}; i = 1, 2, 3, \cdots, n_x; j = 1, 2, 3, \cdots, n_y$

If the sample sizes are not too large these frequencies can readily be obtained by the tabulation of the $n_x n_y$ values of U_{ij} in an $n_x \times n_y$ frequency table by tallying the number of $1s$, 0s and -1s. In general, however, for all sample sizes these frequencies can alternatively be obtained using the sample data of Table 1 as follows.

Let f_{lx}^+ be the number of times a randomly selected subject from population X in condition or level of seriousness l is at a higher, worse or more serious level than subjects at all levels of seriousness in population Y. Also, let f_{lx}^0 be the number of subjects in population X that are in the same level of seriousness l as subjects in population Y; and let f_{lx}^- be the number of subjects in population X at condition l that are in a less, better or lower level of seriousness than subjects at all levels of seriousness in population Y.

Then,

$$f_{lx}^+ = \sum_{h=1}^{l-1} f_{lx} f_{hy};\ f_{lx}^0 = f_{lx} f_{hy};\ f_{lx}^- = \sum_{h=l+1}^{k} f_{lx} f_{hy} \quad (9)$$

Hence, the probabilities that a randomly selected subject from population X at condition or level of seriousness $'l'$ is on the average at a more serious (worse or higher), the same equal) or less serious (better or lower) condition than subjects from population Y in all levels of seriousness are estimated as respectively.

$$\hat{\pi}_{lx}^+ = \frac{f_{lx}^+}{f_{lx} \cdot n_y} = \frac{\sum_{h=1}^{l-1} f_{lx} f_{hy}}{f_{lx} \cdot n_y} = \frac{\sum_{h=1}^{l-1} f_{hy}}{n_y} = \sum_{h=1}^{l-1} \hat{\pi}_{hy} \quad (10)$$

The probability that a randomly selected subject from population X at the l^{th} level of seriousness is on the average at the same level of seriousness or condition $'l'$ as a randomly selected subject from population Y is

$$\hat{\pi}_{lx}^{0} = \frac{f_{lx}^{0}}{f_{lx}.n_y} = \frac{f_{ly}}{n_y} = \hat{\pi}_{l_y} \qquad (11)$$

The probability that a randomly selected subject from population X at condition or level of seriousness $'l'$ is on the average at a less serious (better or lower) condition than subjects from population Y at all levels of seriousness or condition is

$$\hat{\pi}_{lx}^{-} = \frac{f_{lx}^{-}}{f_{lx}.n_y} = \frac{\sum_{h=l+1}^{k} f_{l_x} f_{h_y}}{f_{l_x}.n_y} = \frac{\sum_{h=l+1}^{k} f_{h_y}}{n_y} = \sum_{h=l+1}^{k} \hat{\pi}_{h_y} \qquad (12)$$

for $l = 1, 2, \cdots, k$.

Hence π_x^+, π_x^0 and π_x^- may also be estimated as

$$\hat{\pi}_x^{+} = \frac{\sum_{l=1}^{k} f_{l_x} \hat{\pi}_{lx}^{+}}{n_x};\ \hat{\pi}_x^{0} = \frac{\sum_{l=1}^{k} f_{l_x}.\hat{\pi}_{lx}^{0}}{n_x};\ \hat{\pi}_x^{-} = \frac{\sum_{l=1}^{k} f_{l_x} \hat{\pi}_{lx}^{-}}{n_x} \qquad (13)$$

The corresponding frequencies and proportions for population Y when populations X and Y are respectively the reference and comparison populations can be similarly obtained. Thus,

$$f_{h_y}^{+} = \sum_{l=1}^{h-1} f_{h_y} f_{l_x};\ f_{h_y}^{0} = f_{h_y} f_{h_x};\ f_{h_y}^{-} = \sum_{l=h+1}^{k} f_{h_y} f_{l_x} \qquad (14)$$

And

$$\hat{\pi}_{h_y}^{+} = \sum_{l=1}^{h-1} \frac{f_{l_x}}{n_x} = \sum_{l=1}^{h-1} \hat{\pi}_{l_x};\ \hat{\pi}_{h_y}^{0} = \frac{f_{l_x}}{n_x} = \hat{\pi}_{lx};\ \hat{\pi}_{h_y}^{-} = \sum_{l=h+1}^{k} \frac{f_{l_x}}{n_x} = \sum_{l=h+1}^{k} \hat{\pi}_{l_x} \qquad (15)$$

So that

$$\hat{\pi}_y^{+} = \sum_{h=1}^{k} \frac{f_{h_y}.\hat{\pi}_{h_y}^{+}}{n_x};\ \hat{\pi}_y^{0} = \frac{f_{h_y}.\hat{\pi}_{h_y}^{0}}{n_x};\ \hat{\pi}_y^{-} = \sum_{h=1}^{k} \frac{f_{h_y} \hat{\pi}_{h_y}^{-}}{n_x} \qquad (16)$$

These estimates provide additional useful information that are not readily available when the original ridit analysis or its modified version is used to analyze the data which is an added advantage of the proposed simplified method over and above its earlier versions that may be of further guidance in the implementation of any desired remedial interventionist measures when resources are scarce.

Now from Equation 9 we have that the total number of subjects from population X that are in more serious (worse, higher), the same (equal) or less serious (better, lower) condition than subjects from population Y are respectively

$$f_x^{+} = \sum_{l=1}^{k} f_{l_x}^{+};\ f_x^{0} = \sum_{l=1}^{k} f_{lx}^{0};\ f_x^{-} = \sum_{l=1}^{k} f_{lx}^{-} \qquad (17)$$

The corresponding estimated probabilities are respectively

$$\hat{\pi}_x^{+} = \frac{f_x^{+}}{n_x.n_y};\ \hat{\pi}_x^{0} = \frac{f_x^{0}}{n_x.n_y};\ \hat{\pi}_x^{-} = \frac{f_x^{-}}{n_x n_y} \qquad (18)$$

Also from Equations 6 and 16 we have that

$$W_x = n_x n_y(\hat{\pi}_x^+ - \hat{\pi}_x^-) = f_x^+ - f_x^- \quad (19)$$

A null hypothesis that may be of interest in ridit analysis is that subjects from one of the sampled populations X, say, is in at least the same level of seriousness or condition as subjects from the other population Y, say. That is a null hypothesis of interest would be

$$H_0{:}\pi_x^+ = \pi_x^- + \gamma_0 \text{ or } H_0{:}\pi_x^+ - \pi_x^- = \gamma_0 \text{ versus } H_1{:}\pi_x^+ - \pi_x^- \neq \gamma_0 (0 \leq \gamma_0 \leq 1) \quad (20)$$

Now the sample estimate of the variance of W_x is from Equations 7, 18 and 19

$$Var(W_x) = n_x n_y(\hat{\pi}_x^+ + \hat{\pi}_x^- - (\hat{\pi}_x^+ - \hat{\pi}_x^-)^2) \quad (21)$$

Under the null hypothesis of Equation 20, the test statistic

$$\chi^2 = \frac{(W_x - n_x n_y \gamma_0)^2}{Var(W_x)} = \frac{(W_x - n_x n_y \gamma_0)^2}{n_x n_y(\hat{\pi}_x^+ + \hat{\pi}_x^- - (\hat{\pi}_x^+ - \hat{\pi}_x^-)^2)} \quad (22)$$

has approximately the Chi-square distribution with 1 degree of freedom and may be used to test the null hypothesis. The null hypothesis is rejected at the α level of significance if

$$\chi^2 \geq \chi^2_{1-\alpha;1} \quad (23)$$

Otherwise H_0 is accepted.

The proposed Simplified Approach to Ridit Analysis yields the same estimated values $\hat{\pi}_x^+, \hat{\pi}_x^0$ and $\hat{\pi}_x^-$ as its earlier Modified Version under reference. The main difference between the two is that the former test statistic here designated as W_m has a higher variance. Thus the simplified test statistic W_x is relatively more efficient and hence likely to be more powerful than the modified test statistic, W_m. To show this, we have as reported by [3] and [4] that the estimated variance of the modified ridit analysis, W_m as developed by [5] and [6] is

$$Var(W_m) = \frac{n_x n_y(n_x + n_y + 1)}{3}\left(1 - \frac{\sum_{l=1}^{k}(t_{lxy}^3 - t_{lxy})}{(n_x + n_y)^3 - (n_x + n_y)}\right) \quad (24)$$

which in all cases where n_x and n_y are fairly large $(n_x, n_y \geq 8)$ can with little loss of information be approximated as

$$Var(W_m) = \frac{n_x n_y(n_x + n_y + 1)}{3} \quad (25)$$

Hence the relative efficiency of the proposed simplified test statistic, W_x, to the earlier modified test statistic, W_m, is about

$$RE(W_x; W_m) = \frac{Var(W_m)}{Var(W_x)} = \frac{n_x + n_y + 1}{3(\hat{\pi}_x^+ + \hat{\pi}_x^- - (\hat{\pi}_x^+ - \hat{\pi}_x^-)^2)} \geq \frac{n_x + n_y + 1}{3(1 - \hat{\pi}_x^0)}$$

Since $(\hat{\pi}_x^+ - \hat{\pi}_x^-)^2 \geq 0$ and $\hat{\pi}_x^+ + \hat{\pi}_x^- = 1 - \hat{\pi}_x^0$ from Equation 3

Hence,

$$RE(W_x;\, W_m) \geq 1 \quad (26)$$

for all $n_x, n_y \geq 1$ and $\hat{\pi}_x^0 \geq 0$ showing that the proposed simplified test statistic is at least as efficient and powerful as the earlier modified test statistic, W_m, under reference.

Finally, the only effect of interchanging the roles of Y (reference population) and X (comparison population) is that π_x^+ interchanges with π_y^- and π_x^- interchanges with π_y^+. The variance remains unaffected. Hence, like the earlier modified method, the results obtained using the proposed simplified ridit analysis is unaffected by which of the two populations is used as the reference and which is used as the comparison population or by the sizes of the samples drawn and used from these populations.

Illustrative Examples

Here is to illustrate the proposed simplified method both when the sample sizes are relatively small $(n_x, n_y \leq 30)$ with the following examples.

Example one: A public health worker measured the level of concentration of a certain chemical in the blood stream of random sample of 18 male and 25 female workers in a large filling station from very low (scored -2), through normal (scored 0) to very high (scored 2) obtaining the following results.

Male: 2, 1, 1, 2, 2, 1, 0, 2, 2, -2, -1, 1, 0, -2, -2, -2, -1, 0.

Female: 2, -1, 0, -2, -1, 0, -2, -1, 1, 2, 0, -2, 1, 0, -1, -1, -2, -1, -1, -2, 2, -1, 2, 0, 2.

Research interest here is to determine whether the population of male and female filling station service workers differ in their levels of the concentration of the chemical in their blood stream. To use the proposed method to analyze the data, we may apply Equation 1 to the male and female scores on data in Example 1. The results are presented in a 25×18 frequency table (Table2) where Y and X are treated as the reference and comparison populations respectively. From Table 2 we have that $f_x^+ = 145$; $f_x^0 = 84$; $f_x^- = 221$ and from Equation (19) $W_x = f_x^+ - f_x^- = -76$. Also from Equation (18)

$$\hat{\pi}_x^+ = \frac{f_x^+}{n_x n_y} = \frac{145}{450} = 0.322$$

$$\hat{\pi}_x^0 = \frac{f_x^0}{n_x n_y} = \frac{84}{450} = 0.187$$

and

$$\hat{\pi}_x^- = \frac{f_x^-}{n_x n_y} = \frac{221}{450} = 0.491$$

From Table 2 and Equation 9 if male population Y is used as the reference group then we have that the frequencies for female (X), the comparison population, are

$$f_{2x}^+ = 8(4) = 32;\ f_{3x}^+ = 5(4+2) = 30;\ f_{4x}^+ = 2(4+2+3) = 18;\ f_{5x}^+ = 5(4+2+3+4) = 65;$$

$$f_{1x}^0 = 5(4) = 20;\ f_{2x}^0 = 8(2) = 16;\ f_{3x}^0 = 5(3) = 15;\ f_{4x}^0 = 2(4) = 8;\ f_{5x}^0 = 5(5) = 25$$

$$f_{1x}^- = 5(2+3+4+5) = 70;\ f_{2x}^- = 8(3+4+5) = 96;\ f_{3x}^- = 5(4+5) = 45;\ f_{4x}^- = 2(5) = 10.$$

These values which are summarized in Table 3A are used in Equations 10 – 13 to estimate the corresponding probabilities for X.

Similarly, if female population is used as the reference population then the corresponding frequencies and estimated probabilities are calculated for population, Y $(Male, now\ the\ comparison\ population)$ using Equations 14 and 16. The results are summarized in Table 3B.

In Table 3 (A and B), it could be observed that $\hat{\pi}_x^+ = \hat{\pi}_y^-$ and $\hat{\pi}_x^- = \hat{\pi}_y^+$.

Now from Equation 21,

$$Var(W_x) = n_x n_y (\hat{\pi}_x^+ \hat{\pi}_x^- - (\hat{\pi}_x^+ - \hat{\pi}_x^-)^2) = 450(0.322 * 0.491 - (0.322 - 0.491)^2) = 352.800$$

Hence, the test statistic for the null hypothesis of Equation 20 with $\gamma_0 = 0$ is from Equation 22 obtained as

$$\chi^2 = \frac{W_x^2}{Var(W_x)} = \frac{-76^2}{352.800} = \frac{5776}{352.800} = 16.372\ (P - value = 0.000)$$

which with one degree of freedom is highly statistically significant indicating that female workers at the filling station on the average have lower levels of concentration of the chemical in their blood stream than male workers.

The illustration above showed the application of the proposed simplified procedure to relatively small sample data presented as raw scores. The frequencies f_x^+, f_x^0, and f_x^- are then simply obtained by tallying the numbers of 1s, 0s and -1s in the $n_x n_y$ values of these number u_{ij} cross-classified in an $n_x \times n_x$ frequency table.

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

The proposed method can also be applied to any already summarized sample data presented in the form of Table 1 as illustrated with the following example.

Example 2: The following (Table 4) are data on the severity of road accidents sustained by random samples of motor vehicle drivers by Age of the driver during the year in a certain community.

Research interest here is to determine whether motor vehicle drivers of different ages experience the same level of severity of injury when involved in road accidents.

If the age group ≥ 25 (Y) and age group < 25 (X) are used as the reference and comparison populations respectively, then Equation 9 may be used to calculate the frequencies corresponding to X. Similar analyses are also performed when X and Y are respectively the reference and comparison population. The results are shown in Tables 5 A and B.

From Table 5, as expected, $\hat{\pi}_x^+ = \hat{\pi}_y^-$ and $\hat{\pi}_x^- = \hat{\pi}_y^+$. From Equation 19, $W_x = f_x^+ - f_x^- = 23630 - 28543 = -4913$. Also from Equation 21,

$$Var(W_x) = 103 \times 685(0.335 + 0.405 - (0{,}335 - 0.405)^2) = 51857.925$$

To test the null hypothesis of Equation 20 with $\gamma_0 = 0$, we have from Equation 22 that

$$\chi^2 = \frac{(-4913)^2}{51857.925} = 465.456$$

with a P −value of 0.000 which with 1 degree of freedom is highly statistically significant indicating that younger drivers are less likely than older drivers to experience serious injuries when involved in motor vehicle accidents in the sampled community.

To compare the results obtained with the newly developed simplified method and what would have been obtained if the data in Examples 1 and 2 were analyzed using the Modified Ridit Analysis, as already pointed out, the estimated proportions are the same no matter which of the two methods is used to analyze the data. However, the estimated variance of W_m for the earlier modified method when applied to Example 1 is from Equation 24

$$Var(W_m) = \frac{(25 \times 18)(25 + 18 + 1)}{3}\left(1 - \frac{(720 + 990 + 504 + 210 + 990)}{43^3 - 43}\right) = 6316.200$$

So that the test statistic for the null hypothesis of Equation 20 with $\gamma_0 = 0$ and $W_x = -76$ is from Equation 22

$$\chi^2 = \frac{-76^2}{6316.200} = 0.914$$

which with 1 degree of freedom is now not statistically significant and contradicts result obtained when the simplified alternative procedure was used.

Similarly the estimated variance of the data of Example 2 as calculated from Equation 24 is

$$Var(W_m) = 18555958.51$$

Hence, the corresponding chi – square statistic for the null hypothesis of Equation 20 with $\gamma_0 = 0$ and $W_x = -4913$ is from Equation 22

$$\chi^2 = \frac{-4913^2}{18555958.51} = 1.301$$

with a P-value of 0.1619 which with 1 degree of freedom is not statistically significant, a result that is inconsistent with the result obtained with the simplified procedure.

These contradictory results are due to the fact that the test statistic W_m generally has higher variance and hence is relatively less efficient than W_x. As a matter of fact, the ratios of the estimated variances of W_m to W_x are 17.90 and 357.82 for Examples 1 and 2 respectively. The results would tend to lead to an acceptance of a false null hypothesis (Type II Error) more frequently using W_m than W_x.

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

Also, to compare the present method with what would have been obtained if the original method of Ridit Analysis was used to analyze the data, the mean ridit for X and its associated variance are respectively

$$\underline{r}_x = \hat{\pi}_x^+ + \frac{1}{2}\hat{\pi}_x^-; Var(\underline{r}_x) = \frac{1}{12n_x} \quad (27)$$

Hence, for the data of Example 1,

$$\underline{r}_x = 0.322 + 0.5(0.187) = 0.416$$

$$Var(\underline{r}_x) = \frac{1}{12(18)} = 0.005$$

The corresponding chi – square test statistic is

$$\chi^2 = \frac{(\underline{r}_x - 0.5)^2}{Var(\underline{r}_x)} = \frac{(0.416 - 0.5)^2}{0.005} = 1.400$$

with a P – value of 0.1570 and 1 degree of freedom, the test is not significant.

With respect to Equation 2, if Y and X are used as the reference and comparison populations respectively then

$$\underline{r}_x = 0.335 + \frac{1}{2}(0.260) = 0.465$$

with a variance,

$$Var(\underline{r}_x) = \frac{1}{12(103)} = 0.0008$$

Hence, the corresponding chi – square test statistic is

$$\chi^2 = 1236(0.465 - 0.5)^2 = 1.236$$

with a P –value of 0.1645 the test is not statistically significant.

On the other hand X and Y are used as the reference and the comparison populations respectively, then

$$\underline{r}_y = 0.405 + \frac{1}{2}(0.260) = 0.535$$

with estimated variance,

$$Var(\underline{r}_y) = \frac{1}{12(685)} = \frac{1}{8220}$$

Hence, the chi – square test statistic becomes

$$\chi^2 = 8220(0.535 - 0.50)^2 = 8.22$$

with a P –value of 0.0032 and degree of freedom of 1, the test is statistically significant.

DISCUSSION

The inconsistency of results clearly illustrates one of the problems with the original method of ridit analysis (Bross, 1958) which often results in different conclusions depending on which of the sampled populations is used as the reference and which is used as the comparison population. This problem arises especially when the sample sizes are very disparate, as in Example 2 where the ratio of the samples drawn from populations X and Y is about 1:7. These problems are not encountered with the present simplified method of ridit analysis since the results are independent of the sample sizes and of which of the sampled populations is used as the reference and which as the comparison population. The estimated variance of the test statistic calculated using the simplified approach is also smaller than that of the modified test statistic and would hence lead to a more reliable conclusion.

SUMMARY AND CONCLUSION

This paper has developed a simplified and alternative method of ridit analysis comparable to, but easier to use than, the modified method developed by Oyeka [3] and others. The developed method unlike the earlier methods enables the separate estimation of the probabilities that a randomly selected subject at a specific level of seriousness or condition in one of the sampled populations is at a more (worse, higher), the same (equal) or less (better, lower) level of seriousness or condition than subjects from the other population at all levels of seriousness or condition which are additional useful information for policy purposes.

The method is shown to be relatively more efficient and hence more powerful than the earlier modified method which is itself more efficient than the original ridit analysis. The method is shown to be applicable to both small and large samples and this was illustrate with data.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

References

1. Bross, I. J. (1958). How to Use Ridit Analysis. *Biometrics, 14*(1), 18 - 38. Retrieved January 24, 2020

2. Fleiss, J. L., Chilton, N. W., & Wallenstein, S. (1979). Ridit analysis in dental clinical studies. *Journal od Dental Research, 58*, 2080 - 2084.

3. Oyeka, I. C. (1992). A Modified Ridit Analysis. *Journal of Nigerian Statistical Association, 8*(1), 45 - 69.

4. Oyeka, I. C., & Ebuh, G. U. (2013, November 6). A Comparative Study between Ridit and Modified Ridit Analysis. *American Journal of Theoretical and Applied Statistics, 2*(6), 248 - 254. Retrieved January 24, 2020

5. Hajek, J. (1969): Nonparametric Statistics. Holden-Day, San Francisco.

6. Conover, W. J. (1973): Rank Tests for One Sample, Two Samples and K Samples without the Assumptions of a Continuous Distribution Function. Annals of Statistics 1. Pgs 1105 – 1125.

TABLES

TABLE 1: PRESENTATION OF SAMPLE DATA FOR USE WITH RIDIT ANALYSIS

	Sampled Population		
Level of Seriousness or Group ( ) C_l	X	Y	Total
	f_{lx}	f_{hy}	t_{lxy}
C_1	f_{1x}	f_{1y}	t_{1xy}
C_2	f_{2x}	f_{2y}	t_{2xy}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_k	f_{kx}	f_{ky}	t_{kxy}
Total	n_x	n_y	t_{xy}

Academic Journal of Current Research
Vol.9, No.9; September-2022;
ISSN (2343 – 403X);
p –ISSN 3244 – 5621
Impact factor: 6.37

Table 2: Values of u_{ij} for the data of Example 1

FEMALE	MALE																					
	2	1	1	2	2	1	0	2	2	-2	-1	1	0	-2	-2	-2	-1	0	f_{Ix}^{+}	f_{lx}^{0}	f_{lx}^{-}	n_y
2	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	13	5	0	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
0	-1	-1	-1	-1	-1	-1	0	-1	-1	1	1	-1	0	1	1	1	1	0	6	3	9	18
-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	-1	-1	0	4	14	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
0	-1	-1	-1	-1	-1	-1	0	-1	-1	1	1	-1	0	1	1	1	1	0	6	3	9	18
-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	-1	-1	0	4	14	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
1	-1	0	0	-1	-1	0	1	-1	-1	1	1	0	1	1	1	1	1	1	9	4	5	18
2	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	13	5	0	18
0	-1	-1	-1	-1	-1	-1	0	-1	-1	1	1	-1	0	1	1	1	1	0	6	3	9	18
-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	-1	-1	0	4	14	`18
1	-1	0	0	-1	-1	0	1	-1	-1	1	1	0	1	1	1	1	1	1	9	4	5	18
0	-1	-1	-1	-1	-1	-1	0	-1	-1	1	1	-1	0	1	1	1	1	0	6	3	9	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	-1	-1	0	4	14	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	0	0	0	-1	-1	0	4	14	18
2	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	13	5	0	18
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1	-1	1	1	1	0	-1	4	2	12	18
2	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	13	5	0	18
0	-1	-1	-1	-1	-1	-1	0	-1	-1	1	1	-1	0	1	1	1	1	0	6	3	9	18
2	0	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	13	5	0	18
f_{jy}^{+}	0	5	5	0	0	5	7	0	0	20	12	5	7	20	20	20	12	7	**145**	**84**	**221**	**450**
f_{jy}^{0}	5	2	2	5	5	2	5	5	5	5	8	2	5	5	5	5	8	5	**84**			
f_{jy}^{-}	20	18	18	20	20	18	13	20	20	0	5	18	13	0	0	0	5	13	**221**			
n_x	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	**450**			

Table 3: Summary of Results using the proposed method with data of Example 1

Table 3A: Male, Y, as reference population, Female, X, as comparison population							
Level of Seriousness	**Frequencies**			**Total**	**Proportions**		
	f_{lx}^{+}	f_{lx}^{0}	f_{lx}^{-}		$\hat{\pi}_{lx}^{+}$	$\hat{\pi}_{lx}^{0}$	$\hat{\pi}_{lx}^{-}$
(-2; Very Low)	0	20	70	90	0	0.222	0.778
(-1; Low)	32	16	96	144	0.222	0.111	0.607
0 (Normal)	30	15	45	90	0.333	0.167	0.5
1 (High)	18	8	10	36	0.5	0.222	0.278
2 (Very High)	65	25	0	90	0.722	0.278	0
Total	145	84	221	450	0.322	0.187	0.491

Table 3B: Female, X, as reference population, Male, Y, as comparison population							
Level of Seriousness	**Frequencies**			**Total**	**Proportions**		
	f_{ly}^{+}	f_{ly}^{0}	f_{ly}^{-}		$\hat{\pi}_{ly}^{+}$	$\hat{\pi}_{ly}^{0}$	$\hat{\pi}_{ly}^{-}$
(-2; Very Low)	0	20	80	100	0	0.2	0.8
(-1; Low)	10	16	24	50	0.2	0.32	0.48
0 (Normal)	39	15	21	75	0.52	0.2	0.28
1 (High)	72	8	20	100	0.72	0.08	0.2
2 (Very High)	100	25	0	125	0.8	0.2	0
Total	221	84	145	450	0.491	0.187	0.322

Table 4: Data on the severity of road accidents sustained by driver by Age of driver.

Severity level of Accident	Age of Driver (Years)		
	$< 25\ (X)$	$\geq 25\ (Y)$	Total
	f_x	f_y	t_{xy}
None	35	56	91
Moderate	30	463	493
Severe	18	96	114
Critical	6	22	28
Fatal	14	48	62
Total	103	685	788

Table 5: Summary of Results using the Proposed Method with Data of Example 2

Table 5A: Age $\geq 25\ (Y)$ **as Comparison Population**

Level of severity of Accident	Frequencies			Total	Proportions			Total
	f_{ly}^{+}	f_{ly}^{0}	f_{ly}^{-}		$\hat{\pi}_{ly}^{+}$	$\hat{\pi}_{ly}^{+}$	$\hat{\pi}_{ly}^{0}$	
None	0	1960	22015	23975	0	0.082	0.918	1
Moderate	1680	13890	4980	20550	0.082	0.676	0.242	1
Severe	9342	1728	1260	12330	0.758	0.14	0.102	1
Critical	3690	132	288	4110	0.898	0.032	0.07	1
Fatal	8918	672	0	9590	0.93	0.07	0	1
Total	23630	18382	28543	70555	0.335	0.26	0.405	

Table 5: Summary of Results using the Proposed Method with Data of Example 2

Table 5B: Age $< 25\ (X)$ **as Comparison Population Age**

Level of severity of Accident	Frequencies			Total	Proportions			Total
	f_{lx}^{+}	f_{x}^{0}	f_{lx}^{-}		$\hat{\pi}_{lx}^{+}$	$\hat{\pi}_{lx}^{+}$	$\hat{\pi}_{lx}^{0}$	
None	0	1960	3808	5768	0	0.34	0.66	1
Moderate	16205	13890	17594	47689	0.34	0.291	0.389	1
Severe	6240	1728	1920	9888	0.631	0.175	0.194	1
Critical	1826	132	308	2266	0.806	0.058	0.136	1
Fatal	4272	672	0	4944	0.864	0.136	0	1
Total	28543	18382	23630	70555	0.405	0.26	0.335	