File Copy

UM-HSRI-SA-74-10

A Note on Ridit Analysis

Jairus D. Flora

Technical Report
November 1974

Distributed to
Motor Vehicle
Manufacturers Association

Highway Safety Research Institute
The University of Michigan
Ann Arbor, Michigan

A NOTE ON RIDIT ANALYSIS

Jairus D. Flora, Jr.

Department of Biostatistics, School of Public Health
and Highway Safety Research Institute,
University of Michigan, Ann Arbor, Michigan 48104, U.S.A.

SUMMARY

Ridit analysis as described by Bross [1958] may depend on an arbitrary choice of a reference population. Further, in the frequently occurring situation in which two populations are sampled for comparison, the ridit analysis ignores the essential two-sample nature of the problem. Thus, although the mean ridit is a useful point estimate of $P(X \leq Y)$, tests of hypotheses or calculation of confidence intervals may be misleading. A modified procedure is proposed which removes the difficulties in testing or forming confidence intervals, and results in a more informative summary of the data. The procedure is illustrated on a set of accident injury data and compared with the ridit analysis. Some simulation comparisons are also reported.

1. INTRODUCTION

Bross [1958] has suggested the use of Ridit analysis for data which are ordered, but are not on an interval scale, such as injury categories. The procedure is as follows. From a reference population with the same categories (of injury, say) one determines a "ridit" or score for each category. This score for each category is the percentile rank of an item in the reference population and is equal to the number of items in all lower categories plus one-half the number of items in the subject category, all divided by the population size. Once the ridits for each category have been determined, they are taken as values of a dependent variable for the other (comparison) groups and the usual normal distribution family of statistics is applied (e.g., means, standard deviation, etc.). The mean ridits calculated in this way will be approximately normal for reasonable sample sizes.

The interpretation of the mean ridit for the comparison group is as follows. If an item, Y, is selected at random from the reference population and an item X, is selected at random from the comparison group, then the mean ridit is an estimate of $P(Y \leq X)$, that is, of the probability that Y is no more seriously injured than X. The mean ridit calculated for the reference population will always be 0.5 by its definition. The variance of the ridit scores in the reference population will depend on the shape of the distribution, being at most 1/4, about 1/12 for approximately equal numbers in the categories, and quite a bit less for extremely skewed distributions.

The maximum variance of the mean ridit in the comparison group is 1/4m. This fact can be used to give a conservative test of whether the mean ridit differs from that of the reference population based on

$$t' = (\bar{X} - 0.5)/\sqrt{1/4m}$$

$$= 2\sqrt{m}(\bar{X} - 0.5)$$

This test is usually ultra conservative and the t test based on

$$t = \sqrt{m}(\bar{X} - 0.5)/s$$

is usually used. Typical values for s^2 for accident injury data have been on the order of .17 as compared to .5 for the upper bound and .28 for the uniform approximation.

Implicit in Bross's work is the assumption that the reference group is a population. He mentions the difficulty in selecting an appropriate reference group, but does not explicitly suggest an appropriate procedure when either of two groups to be compared might serve as a reference group. In applications of ridit analysis to date the reference group has been the same general size as the comparison group and is generally not a population. If the reference group is much larger than the comparison group, Bross's procedure is still appropriate; however, this is typically not the case.

The difficulty is caused by the problem of determining an appropriate standard deviation to use in the denominator of the t statistic. Interchanging the roles of the reference and comparison groups merely interchanges the X and Y. The mean ridit still estimates a useful probability. However, the choice of which population is to be the reference group affects the value of s^2 and hence the result of a test of hypothesis or the length of a confidence interval. Further, if both groups are regarded as samples from their respective populations, an additional source of variability is introduced; the ridit scores are subject to variation themselves. Monte Carlo investigations (presented in Section 4) indicate that this variability makes all of the estimates of the standard deviation suggested by Bross inappropriate.

In the following section we outline a procedure which is similar to ridit analysis in interpretation, but which makes the explicit assumption that the data are to be regarded as samples from both populations. This procedure is based on the Wilcoxon-Mann-Whitney test. Recent work by Conover [1972] provides a theoretical basis. This procedure and ridit analysis are compared in Section 3 using accident injury data from an investigation of side-door-beam effectiveness. Section 4 presents some simulation results.

2. THE PROPOSED PROCEDURE

Let one group of observations be represented by Y_i (Bross's reference group) and the other group by X_j (the comparison group).

Define

$$D_{ij} = \begin{cases} +1 \text{ if } Y_i > X_j \\ 0 \text{ if } Y_i = X_j \\ -1 \text{ if } Y_i < X_j \end{cases} \tag{1}$$

for $i=1, \cdots, n$; $j=1, \cdots, m$. Here n is the size of the "Y" sample, m is the size of the "X" sample. The test is based on the statistic

$$W = \sum_{i=1}^{n} \sum_{j=1}^{m} D_{ij}. \tag{2}$$

Let

$$\begin{aligned} \pi^+ &= P[Y>X] \\ \pi^0 &= P[Y = X] \\ \pi^- &= P[Y<X] \end{aligned} \tag{3}$$

denote the probability that a member of the reference group is worse off, as well off, or better off, respectively than a member of the comparison group. (Bross's mean ridit estimates $\pi^+ + 1/2\pi^0$.) We then have

$$E[W] = mn(\pi^+ - \pi^-), \tag{4}$$

so that W/mn estimates $\pi^+ - \pi^-$.

With very little additional effort we can get estimates of π^+, π^-, and π^0 individually. The data will be described by the following table.

TABLE 1

DATA FORMAT

Ordered categories:	C_1	C_2	$\cdots$	C_k	Total
Group Y (reference)	f_1	f_2	$\cdots$	f_k	n
Group X (comparison)	g_1	g_2	$\cdots$	g_k	m
Total	τ_1	τ_2	$\cdots$	τ_k	N = m+n

Thus the k categories are denoted $C_1, \cdots, C_k$, say, ranging from least to most severe injury. The observed frequencies of the Y group are f_i and of the X group are g_i and the totals for each category are τ_i.

Then the variance of W is given by

$$\mathrm{Var}\ (W) = \frac{mn(N+1)}{3} \left[1 - \frac{\sum_{i=1}^{k} (\tau_i^3 - \tau_i)}{N^3 - N} \right] \tag{5}$$

Further, W is approximately normally distributed for large m and n. This follows from Theorem 4.1 of Conover [1972] or from Theorem 29C of Hājek [1969]. The goodness of the normal approximation depends on the sample sizes and also the τ_i, but for sample sizes usually encountered with this sort of data, the approximation should be satisfactory. From data in the form of Table 1, W is most easily computed by the formula

$$W = \sum_{i=2}^{k} f_i \left(\sum_{j=1}^{i-1} g_j \right) - \sum_{i=1}^{k-1} f_i \left(\sum_{j=i+1}^{k} g_j \right). \tag{6}$$

The quantity

$$\hat{\Pi}^o = \left(\sum_{i=1}^{k} f_i\ g_i \right) / mn \tag{7}$$

serves as an estimate of Π^o. Note that

$$1 - \Pi^o = \Pi^+ + \Pi^-.$$

Thus

$$\hat{\Pi}^+ = 1/2(1 - \hat{\Pi}^o + W/mn). \tag{8}$$

If the two groups came from the same population $\Pi^+ = \Pi^-$ so the expected value of W is zero. This fact is used to test the hypothesis that the two populations being sampled are equivalent with respect to injury severity. For example, if the addition of a side door beam does not change the chance of injury in a side collision, the probability that a person with a side door beam will be more seriously injured than a person without a side beam would be the same as the probability that he will be less seriously injured with than without.

Thus the test statistic for the hypothesis that the two populations are the same ($\Pi^+ = \Pi^-$) is based on

$$Z = \frac{W}{\sqrt{Var(W)}}$$

The value of Z thus calculated is compared to the standard normal table to obtain the significance.

A confidence interval for the difference ($\Pi^+ - \Pi^-$) may be calculated as follows. From the normal table obtain the value $Z_{\alpha/2}$ for a two-sided level of α. The confidence coefficient will be $1 - \alpha$. Then the interval is calculated as

$$\frac{W}{mn} \pm Z_{\alpha/2} \left\{\frac{Var(W)}{m^2n^2}\right\}^{1/2} \qquad (9)$$

If the interval includes zero, the two sided test of the hypothesis that $\Pi^+ = \Pi^-$ is accepted. If the interval does not include zero, this hypothesis is rejected.

3. EXAMPLES AND ILLUSTRATIVE CALCULATIONS

Two examples are worked in detail. The first uses the data in Table 2, which are from a study by Preston and Shortridge [1973] aimed at investigating the effectiveness of a side door beam in reducing injury to occupants in side collesions. The second set of data is from a similar study by McLean [1973] and is presented in Table 3. If the "no injury" category of the second data set were omitted, the two sets would seem to be comparable.

TABLE 2

INJURY SEVERITY BY GROUP
(DRIVERS IN LEFT SIDE IMPACT)

Group	Severity				Total
	Pain	Minor	Carried	Killed	
No Beam	79	59	40	0	178
Beam	15	10	5	0	30
Total	94	69	45	0	208

3.1. A RIDIT ANALYSIS

Following Bross [1958] one calculates the ridit scores for each group (first using the "No Beam" group as the reference group). We have:

"Pain": $\frac{79/2}{(178)} = .2219$

"Minor": $\frac{79 + 59/2}{178} = .6095$

"Carried": $\frac{79 + 59 + 40/2}{178} = .8876$

"Killed": 1.0

(Notice that since there were no fatalities in either group, the "Killed" category does not enter into the calculations.) From the ridit scores we calculate the mean ridit

$$\bar{X} = \sum_{i=1}^{k} f_i \, r_i/m$$

$$= 0.4621,$$

and the standard deviation of $\bar{X}$,

$$S_{\bar{X}} = [\sum_{i=1}^{k} f_i (r_i - \bar{X})^2/m(m - 1)]^{1/2} = .0478$$

from which

$$t = -0.794 \, .$$

This corresponds to a two-sided significance level of 0.43, indicating that the differences are likely due to chance. Thus the ridit analysis estimates that the probability that an individual with the door beam is as severely injured or more severely injured than one without is .462, or in terms of odds, the odds of a serious injury are 1.00 to .8587 in favor of the side door beam.

If the side door beam group were selected as the reference group instead, the ridit scores would be:

"Pain": .250

"Minor": .667

"Carried": .917

giving a mean ridit for the "no side door beam" group of $\bar{X} = .538$. This value is simply one minus the mean ridit obtained using the other group as the reference group. This corresponds to the fact that a complementary probability is being estimated. However the effect on the standard deviation of the mean can be serious. With these data, we calculate

$$S_{\bar{X}} = 0.0205,$$

leading to

$$t = 1.853.$$

The magnitude and direction of the estimated effect are the same, but the change in choice of the reference group has resulted in a two sided significance of 0.064. The inference drawn could easily have changed from no effect to one of protective effect for the side door beams.

2.2. A MODIFIED ANALYSIS

Consider the statistic W. It is easily computed from the data in Table 2 using (6) as

$$\begin{aligned} W &= 59(15) + 40(15 + 10) - 79(10 + 5) - 59(5). \\ &= 405. \end{aligned}$$

Hence

$$(\hat{\Pi}^{+} - \hat{\Pi}^{-}) = W/mn$$
$$= 0.0758.$$

To estimate Π^{O} we use (7):

$$\hat{\Pi}^{O} = [(79)(15) + (59)(10) + (40)(5)]/(30)(178)$$
$$= 0.3699.$$

Using (8) we find $\hat{\Pi}^{+} = .3530$.

Thus,

$$\hat{\Pi}^{+} = .3530$$
$$\hat{\Pi}^{O} = .3699$$
$$\hat{\Pi}^{-} = .2771 .$$

Thus the analysis estimates that if a driver of a car with a side door beam is involved in an accident involving a left side impact, the probability is .2771 that he would have been less severely injured without the beam; .3699 that he would have sustained the same severity injury; and the probability is 0.3530 that he would have been more severely injured without the beam. Notice that

$$\hat{\Pi}^{-} + 1/2\hat{\Pi}^{O} = .4621$$
$$= \bar{X},$$

except for rounding. The difference in analyses thus far is that the probability of sustaining the same injury has been estimated separately.

With some data sets this is important additional information as is seen in the next example.

The variance of W is calculated from (5). It is convienient to calculate the cubes in the last term.:

T	T^3-T
94	830,490
69	328,440
45	91,080
	1,250,010
208	8,989,704.

Thus

$$V(W) = \frac{30(178)(209)}{3}\left[1 - \frac{1,250,010}{8,989,704}\right]$$

$$= \frac{30(178)(209)}{3}(.86109)$$

$$= 320342.7,$$

$$S(W) = 565.988,$$

and

$$Z = .716, \text{ which is not significant.}$$

Using (9), we can calculate a 95% confidence interval for $\Pi^+ - \Pi^-$. The interval becomes (-.132, .284).

If one were to interchange the groups with the W procedure, the only changes are: 1) the sign of W changes, 2) $\hat{\Pi}^+$ and $\hat{\Pi}^-$ are interchanged,

3) the confidence limits are multiplied by minus one. The significance level of the test statistic remains the same.

3.3 A SECOND EXAMPLE

For the second example we use the data from Table 48 of McLean 1973 .

TABLE 3

INJURY SEVERITY BY GROUP: DRIVERS IN LEFT SIDE IMPACTS

Group	Injury Severity					Total
	None	C	B	A	K	
No Beam	731	41	28	72	9	881
Beam	286	14	10	17	2	329
Total	1017	55	38	89	11	1210

These data will be analyzed twice by each method. The first analysis omits the "None" category to make the data comparable to those of Table 2. The second analysis includes all categories to illustrate the effect of inclusion of a large group of uninjureds on the ridit and W statistics. The "No Beam" group is selected as the reference group since that was the standard design at the time the study was done and since there are more cases in that group. Table 4 summarizes the results of the various analyses and includes the results of reversing the role of the reference and comparison groups for the data in Table 3. As before, the level of significance is drastically affected by the change.

Omitting the "None" category, the ridit scores become

C	B	A	K
.1367	.3667	.7000	.9700 ,

from which $\bar{X} = .4517$ and $S_{\bar{X}} = .0481$, yielding $t = -1.176$. Thus the direction of change is the same as in the data of Table 1, indicating a slight advantage for drivers with the side door beam. The difference does not attain significance, however.

Using the modified analysis, one has $W = 624$, from which

$$\hat{\Pi}^{+} - \hat{\Pi} = .0967.$$

The standard deviation is 602.74, leading to $t = 1.04$, again indicating nonsignificance. The estimated probabilities are

$$\hat{\Pi}^{+} = .3859,$$

$$\hat{\Pi}^{-} = .2891,$$

$$\hat{\Pi}^{0} = .3250,$$

corresponding to probabilities of greater, less, or equal severity of injury, respectively, without the side door beam. The results are quite similar to those from Table 1.

If the uninjured category is included, the sample size will increase noticeably. By far, most of the drivers in both groups fell in this category. The ridit scores now become

None	C	B	A	K
.4149	.8530	.8992	.9489	.9449 ,

from which

$$\bar{X} = .4792,$$

and

$$S_{\bar{X}} = .0092,$$

leading to

$$t = -2.27,$$

which indicates significance at the 5% level. Notice that although the mean ridit is closer to one-half, indicating a smaller magnitude of change, the significance has increased due to the increase in sample size. The interpretation would be that the probability of being more severely injured with a beam than without is only .4792 or in terms of odds, the odds are .92 to 1.0 in favor of the driver with the side door beam.

We find that W = 12,109 and the standard deviation of W is 6887.99, leading to Z = 1.76, not significant at the 5% level. Further, W/mn = .0418, and

$$\hat{\Pi}^{+} = .1565,$$

$$\hat{\Pi}^{-} = .1148,$$

$$\hat{\Pi}^{o} = .7285,$$

estimating the probabilities of greater, less, or equal severity injury, respectively, without the side door beam. From these it is evident (in this set of accidents at least) that one is most likely to sustain the same injury with or without a side-door beam. The estimated difference in probabilities of a more or less severe injury is only .0418. Thus the possible effect is quite small, and does not reach statistical significance at the 5% level. The estimation of Π^{+}, Π^{-}, and Π^{o} separately provides more information and aids in determining whether an effect (statistically significant or not) is large enough to be of practical significance.

Table 4 summarizes the results of the analyses on these two sets of data. Some general conclusions are apparent. First of all, the group with

TABLE 4

COMPARISON OF ANALYSES

Data Set	Ridits			Proposed Procedure				
	$\bar{X}$	t	Sig.	$\hat{\Pi}^+$	$\hat{\Pi}^o$	$\hat{\Pi}^-$	Z	Sig.
Table 2 (B)[1]	.538	1.85	.064	.353	.370	.277	.72	.472
Table 2($\bar{B}$)	.462	-.79	.430	.277	.370	.353	-.72	.472
Table 3 (B) (None omitted)	.548	2.16	.030	.386	.325	.289	1.04	.298
Table 3 ($\bar{B}$) (None omitted)	.452	-1.18	.238	.289	.325	.386	-1.04	.298
Table 3 (B)	.521	3.23	.001	.157	.729	.115	1.76	.078
Table 3 ($\bar{B}$)	.479	-2.27	.023	.115	.729	.157	-1.76	.078

[1] The letter in paranthesis denotes the group used as the reference group: B = Beam, $\bar{B}$ = No Beam.

the side door beam is consistently slightly better off than the group without the door beam. The protective effect is probably real, but appears to be quite small. (It is possible, of course that it might be greater for other occupants, say for those situated on the opposite side of the car from the impact. See Preston and Shortridge [1973] or McLean [1973] for comments on this possibility.) In each case, interchanging the reference and comparison groups has a marked effect on the ridit analysis, changing the attained significance drastically. The change merely results in a change in sign for the proposed procedure. The level attained by the ridit procedure appears to be lower than for the proposed procedure, even when the more conservative group is chosen as the reference group. From simulation results to be presented in the next section this appears to be due to the ridit procedure being anti-conservative; that is, its actual α is larger than the nominal value.

4. SOME SIMULATION RESULTS

As mentioned by Bross [1958], selection of the reference group requires considerable care. The examples of the preceeding section emphasize this, showing a rather drastic effect on inference if the smaller group were chosen as the reference group. In the examples presented, the "no beam" group was selected as the reference group. This was natural since at that time side door beams were an inovation and the bulk of cars did not have them. Thus, in addition to representing a change from the status quo, most accidents could be expected to involve cars without side door beams, making the sample size for the reference group larger than for the comparison group. However, this need not have been the case. The sample sizes might easily have been nearly equal. Further, side door beams are now installed in nearly all new cars. Consequently, the "natural" reference group might be different today. Thus it is desirable to have a procedure which is not sensitive to what could be an arbitrary choice of a reference group. The simulations indicate that when both groups represent samples, the choice of the larger sample size as the reference group is not sufficient to guarantee a valid significance level with the ridit procedure, even when the ratio of sample sizes is about 4 to 1. Of course, when one group is a population, the ridit procedure is valid and would be preferable on the basis of power.

The basic problem inherent in the ridit analysis as exemplified above, is that ridit analysis treats the ridit scores as known and only the experimental group as a sample. In most applications both the experimental group and the reference group are represented by samples and the problem is of the "two-sample" type rather than of the "one-sample" type. As a result the standard error of the mean ridit as usually estimated is too small.

To verify this empirically a simulation was conducted. The populations being sampled were taken to consist of five categories, each with probability 0.2. Both the reference group and the experimental group were drawn from this population, so that the null hypothesis that both groups came from the same population is known to hold. Samples of 200 were drawn for each group and the mean ridit and various estimates ($\sqrt{1/4m}$, $\sqrt{1/12m}$, $S/\sqrt{m}$) were calculated for $S_{\bar{X}}$. The experiment was replicated 250 times. It was found that $S/\sqrt{m}$ and $\sqrt{1/12m}$ both underestimated the standard deviation of $\bar{X}$ by about 40%, while $\sqrt{1/4m}$ overestimated the standard deviation of $\bar{X}$ by about 20%. The detailed results are presented in Table 5. The observed Type I error rates for nominal $\alpha = 10\%$, $\alpha = 5\%$ and $\alpha = 1\%$ are also presented. The last line of the table presents the results of using W.

TABLE 5

RESULTS OF THE SIMULATION

Method	Ratio of Standard Errors	Nominal 10%	5%	1%
$S=\sqrt{1/4m}$	.806	0.052	0.020	0.000
$S=\sqrt{1/12m}$	1.397	0.232	0.152	0.064
$S = S_{\bar{X}}$	1.435	0.244	0.156	0.064
S_W	1.007	0.092	0.052	0.008

It can be seen from Table 5 that the use of $\sqrt{1/4m}$ is too conservative, resulting in smaller than nominal probabilities of Type I error. On the other hand the use of either of the other suggestions for the standard error is anti-conservative, resulting in far too large probabilities of Type I error.

This point is further illustrated by two additional simulations. In these, the null hypothesis no longer holds; the experimental group has been shifted by 20%. Hence, comparison of probabilities of Type I errors is not valid. However, the ratio of the empirical standard errors to the estimates still gives an indication of the difficulty. In the first of these simulations, the population described before was shifted to the left by 20% for the experimental group. In the second simulation, the category probabilities were (.6, .15, .1, .05, .1) for the reference group and (.63, .14, .09, .06, .08) for the experimental group, again representing a 20% shift. The sample sizes were 200 each and the number or replications remained at 250. Again, results for W are reported in the last line.

TABLE 6

RESULTS OF SIMULATIONS UNDER ALTERNATIVES

A. Simulation Cell Probabilities: Uniform with a 20% Shift

Estimate for Standard Error	Ratio of Empirical to Estimated Standard Error	Empirical Power for α of .1	.05	.01
$S=\sqrt{1/4m}$	.782	.132	.092	.012
$S=\sqrt{1/12m}$	1.354	...	...	...
$S=S_{\bar{X}}$	1.418	...	...	...
$S=S_W$	.977	.244	.140	.080

B. Simulation Cell Probabilities (.6, .15, .1, .05, .1) and a 20% Shift

Estimate for Standard Error	Ratio of Empirical to Estimated Standard Error	Empirical Power for α of .1	.05	.01
$S=\sqrt{1/4m}$	.714	.060	.024	.000
$S=\sqrt{1/12m}$	1.237	...	...	...
$S=S_{\bar{X}}$	1.466	...	...	...
$S=S_W$	.999	.200	.124	.032

Since the nominal values for α are wrong, power results are reported only for $S=\sqrt{1/4m}$ and for W. The larger power for W results from the fact that the level for W is the nominal α, while the level for the approximation is somewhat smaller, due to its conservative nature.

We conclude that the modified statistic should be preferred to the earlier ridit analysis in the cases where both groups are represented by samples. If the ridit analysis is to be used, the approximation of

$S=\sqrt{1/4m}$ is recommended for the standard deviation. This is conservative; the other suggestions appear to be considerably anticonservative. The use of the sample standard deviation of $\bar{X}$ in testing is valid only when one of the groups is a population, or when the sample sizes are so disparate that one group may be considered a population. In this case, the larger group must be taken as the reference group.

It should be noted that the statistics involved in the comparisons use the same information. The only point at issue is the appropriate standard deviation to use in the normal approximation. Consequently, for the same α level, the asymptotic relative efficiency of each to the other would be one. The modified procedure can easily be extended to the k-sample situation. In this case it would be essentially a Kruskal-Wallis test with many ties. Adaptations of the multiple comparisons methods associated with rank tests should also be relatively simple.

The author wishes to thank Mr. Robert Scott of the Highway Safety Research Institute, University of Michigan for posing the initial question regarding the validity of ridit analysis and Mr. Hank Goloumb, also of HSRI, for programing the simulations.

REFERENCES

Bross, I. D. J. [1958]. "How to Use Ridit Analysis." Biometrics 14, pg. 18-38.

Conover, W. J. [1973]. "Rank Tests For One Sample, Two Sample, and k Samples Without the Assumption of a Continuous Distribution Function." Annals of Statistics 1, pg. 1105-1125.

Hajek, J. [1969]. Nonparametric Statistics. Holden-Day, San Francisco.

McLean, A. J. [1973]. "Collection and Analysis of Collision Data for Determining the Effectiveness of Some Vehicle Systems." University of North Carolina Report No. 7301-C19.

Preston, F. and Shortridge, R. [1973]. "An Evaluation of the Effectiveness of Side-Door Beams Based on Accident Exposure." University of Michigan Report No. UM-HSRI-SA-73-8.