1992-03

Min and max scorings for two-sample ordinal data

Kimeldorf, George; Sampson, Allan R.; Whitaker, Lyn R.

https://hdl.handle.net/10945/41700

Min and Max Scorings for Two-Sample Ordinal Data

GEORGE KIMELDORF, ALLAN R. SAMPSON, and LYN R. WHITAKER*

To analyze two-sample ordinal data, one must often assign some increasing numerical scores to the ordinal categories. The choice of appropriate scores in these types of analyses is often problematic. This article presents a new approach for reporting the results of such analyses. Using techniques of order-restricted inference, we obtain the minimum and maximum of standard two-sample test statistics over *all possible assignments of increasing scores*. If the range of the min and max values does not include the critical value for the test statistics, then we can immediately conclude that the result of the analysis remains the same no matter what choice of increasing scores is used. On the other hand, if the range includes a critical value, the choice of scores used in the analysis must be carefully justified. Numerous examples are given to clarify our approach.

KEY WORDS: Cochran-Armitage procedure; Isotonic regression; Monotone scales; Ordinal data; Scoring; t test; Two-sample problem.

1. INTRODUCTION

Suppose that we have data drawn from two populations or treatments where each observation falls into one of k levels of an ordinal categorization. For instance, in a clinical trial the populations may correspond to active and control treatments, and the ordinal responses may be the physician's evaluation of the patient's change from study start as one of: very improved, moderately improved, no change, slightly deteriorated, or very deteriorated. The typical goal is to assess whether or not there is a difference between the two treatments. A variety of statistical methods are employed in this setting [see Miller (1986, chap. 2), Hettmansperger (1984, chap. 3), and Agresti (1984)]. These include the Wilcoxon–Mann–Whitney test, the Cochran–Armitage procedure, scored t tests, binomial proportions tests where the ordinal categorization is dichotomized (the ordinal dichotomy may be, for example, "improved" versus "not improved"), and also tests based on log-linear models. In using one of these procedures to analyze data, the practicing statistician, wondering if another procedure might produce different results, may, in fact, report a variety of these test results for these two-sample ordinal data sets. Since there is arbitrary scoring in some of these procedures, there conceptually appears to be a limitless set of possible results to report.

In this article we propose a new approach to reporting results of analyses of this type of data. We begin by noting that these previously noted and commonly used test statistics can essentially be viewed as monotone functions of a certain statistic with increasing numerical scores assigned to the categories. We then present a simple solution to the problem of computing the maximum and the minimum of this statistic over *all* possible assignments of increasing scores. These results then lead us to propose reporting both the maximum and minimum values of the corresponding test statistic when analyzing such data. Obviously, when either min and max statistics both produce statistically significant results or both produce insignificant results, inferences are immediate: any increasing scoring produces the same results concerning whether or not the two populations differ. We call this case *nonstraddling*. In the *straddling* case, we become aware that care and interpretability are important in choosing the numeric scores by which we analyze the data. Further discussion concerning this latter case is presented.

In Section 2, we review various two-sample procedures for ordinal data and represent them in terms of a monotone function of a certain correlation based on increasing scores. We introduce the optimization notions in detail in Section 3 and present the computational aspects for these in Section 4. Section 5 offers a brief discussion, and proofs are in the Appendix.

2. TWO–SAMPLE ORDINAL PROCEDURES: A REVIEW

Suppose that the data are drawn from two populations or treatments denoted for convenience as 0 and 1, where the data are ordinal and each observation falls into one of k levels. The levels are denoted as $L_1 < \cdots < L_k$, where $<$ denotes the underlying experimental ordering. It is usual to represent such two-sample data in tabular form as in Table 1, where, for example, m_i is the number of observations from population 0 that fall into level L_i.

To assess whether or not there is a difference between these populations, a variety of simple statistical procedures are currently used in practice. The nonparametric Wilcoxon–Mann–Whitney two-sample test is often used in this setting (see Hettmansperger 1984). When there is a natural ordinal dichotomous grouping of the levels, populations are sometimes compared by computing a χ^2 statistic from the resultant 2×2 table. Another common approach is to assign increasing scores $x_1 \leq \cdots \leq x_k$ ($x_1 \neq x_k$) to the ordinal levels $L_1, \ldots, L_k$, and then use the usual two-sample t test based on these scores. In this t test approach, implicitly assumed is that the sample sizes m and n are "sufficiently

* George Kimeldorf is Professor, Programs in Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75083-0688. Allan R. Sampson is Professor, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260. Lyn R. Whitaker is Assistant Professor, Department of Operations Research, Naval Postgraduate School, Monterey, CA 93940. Allan R. Sampson's research was supported by National Security Agency Grant MDA90490H4036. Reproduction in whole or part is permitted for any purpose of the U.S. government. Lyn R. Whitaker's research was supported in part by the Research Foundation at the Naval Postgraduate School. The authors thank Shiva Gautam for bringing to their attention the necessary results in Bazaraa and Shetty (1979) and for the insight in the use of their results in the proof of Theorem A.3. Moreover, the authors thank Anthony Harris for his many insightful conversations and also Leon Gleser for a beneficial discussion. Additionally, the associate editor and referees provided a number of helpful comments leading to an improved final version of the article.

Journal of the American Statistical Association
March 1992, Vol. 87, No. 417, Theory and Methods

Table 1. Two-Sample Data

		Levels				
		L_1	L_2	...	L_k	Total
Population	0	m_1	m_2	...	m_k	m
	1	n_1	n_2	...	n_k	n
	Total	$m_1 + n_1$	$m_2 + n_2$	...	$m_k + n_k$	N

large" enough to justify the t distribution as a good approximation to the null hypothesis reference distribution. The t statistic computed with arbitrary increasing scores $x_1, \ldots, x_k$ becomes

$$t(x_1, \ldots, x_k) = [(N-2)/N]^{1/2}(mn)^{1/2} \times (A_1 - A_0)\Big/\left[\sum_{i=1}^{k} (m_i + n_i)x_i^2 - mA_0^2 - nA_1^2\right]^{1/2},$$

where $A_0 = \Sigma\, m_i x_i/m$ and $A_1 = \Sigma\, n_i x_i/n$. In the same spirit, the Cochran–Armitage statistic [e.g., Agresti (1990, sec. 4.4.3) or Mantel (1963)] is a function of such scores, namely,

$$C(x_1, \ldots, x_k) = [(N-1)^{1/2} r]^2, \qquad (2.1)$$

where $r = r(x_1, \ldots, x_k)$ is the Pearson correlation coefficient based on the scores $x_1, \ldots, x_k$ and values 0 and 1 assigned to populations, that is

$$(N-1)^{1/2} r(x_1, \ldots, x_n) = [(N-1)/N]^{1/2}(mn)^{1/2} \times (A_1 - A_0)\Big/\left[\sum_{i=1}^{k} (m_i + n_i)x_i^2 - N^{-1}(mA_0 + nA_1)^2\right]^{1/2}.$$

The null hypothesis asymptotic distribution of the C statistic is a χ^2 distribution with 1 df. Often, the Cochran–Armitage test is made into a two-sided test by comparing $(N-1)^{1/2} r$ to a standard normal distribution.

As is well known (e.g., Woodward and Overall 1977, p. 170), the C statistic and the t^2 statistic are nearly identical in value. To see this, note that

$$t^2(x_1, \ldots, x_k) = [(N-2)/(N-1)]^{1/2} \times (C(x_1, \ldots, x_k))/(1 - C(x_1, \ldots, x_k)/(N-1)).$$

Thus for large N, both statistics' values will be approximately the same. No matter the sample size, however, t^2 is a strictly increasing function of C. Most interesting to observe is that

$$t(x_1, \ldots, x_k) = (N-2)^{1/2} \times r(x_1, \ldots, x_k)/(1 - r^2(x_1, \ldots, x_k))^{1/2} \qquad (2.2)$$

is also a strictly increasing function of $r(x_1, \ldots, x_k)$, as is the C statistic itself.

The choice of which scores to employ for either the t statistic or the C statistic can be problematic. See Graubard and Korn (1987) for a nice discussion in this setting about the choice of scores and their effects on the results of the analysis. Some typically used scoring systems in data analyses include: $1, \ldots, k$, which are called uniform scores or equal-spacing scores; $R_1, \ldots, R_k$, which are marginal midrank scores; $R_1/N, \ldots, R_k/N$, which are ridits (e.g., Fleiss 1981, Sec. 9.4); and $R_1/(N+1), \ldots, R_k/(N+1)$, which are called modified ridit scores (Lehman 1975). In fact, in conducting the Cochran–Armitage test, the FREQ procedure of SAS (SAS Institute, Inc. 1985) allows the user the choice of the preceding scores as well as arbitrary user-provided scores.

Interestingly, various choices of scores for the t test or the Cochran–Armitage test reduce those procedures to other related procedures. For example, the assignment of marginal midrank scores essentially reduces the t test to the Wilcoxon–Mann–Whitney two-sample test (e.g., Conover and Iman 1981, sec. 2). Utilization of the scores $x_1 = 0, \ldots, x_j = 0, x_{j+1} = 1, \ldots, x_k = 1$ reduces the C statistic to $[(N-1)/N] \times (\chi^2)$, the statistic for the 2×2 table, where levels $L_1, \ldots, L_j$ are grouped and $L_{j+1}, \ldots, L_k$ are grouped. We note that, for the latter special case, Miller and Siegmund (1982) and Halpern (1982) considered the problem of choosing the scores for dichotomizing the ordinal levels that maximize the χ^2 statistic and studied the effects of optimization on the distribution of the resultant χ^2 statistic.

A standardly used log-linear model in this case (see Agresti 1984) can be parameterized as

$$\log E(m_i) = \mu + \lambda^0 + \lambda_i, \qquad i = 1, \ldots, k,$$

and

$$\log E(n_i) = \mu + \lambda^1 + \lambda_i + \beta x_i, \qquad i = 1, \ldots, k,$$

where the increasing scores x_i $(i = 1, \ldots, k)$ are assigned. It is known [e.g., Agresti (1984, note 2, p. 98) or Agresti (1990, sec. 8.1.6)] that the efficient score statistic for testing whether or not $\beta = 0$ is essentially the t test for correlation with the noted scores. For further discussion of this test, see Agresti (1990, chap. 8) or Agresti, Mehta, and Patel (1990).

To reiterate, the scored t statistics and the C statistic are monotone functions of $r(x_1, \ldots, x_k)$ and are essentially equivalent tests. Furthermore, for particular choices of scores, these tests reduce to the nonparametric Wilcoxon–Mann–Whitney and also to the grouped 2×2 χ^2 test.

3. USING OPTIMIZED STATISTICS

In the general context of ordinal categorical data analysis, Agresti (1984) writes

> sometimes it is not obvious how to assign scores. . . . In such cases it is informative to assign scores a variety of "reasonable" ways to check whether substantive conclusions depend on the actual choice. (p. 97)

This approach to using a variety of scores in analysis appears to be a fairly standard practice among statisticians analyzing data. Rather than computing the t statistic or the C statistic for a variety of scores, we propose that a more basic calculation be made. For the t statistic, the statistician should find both the minimum and the maximum of $t(x_1, \ldots, x_k)$ over all nondegenerate increasing scores. Similarly, when working with the Cochran–Armitage procedure the statistician should compute the minimum and maximum of $C(x_1, \ldots, x_k)$, or of its one-sided version $(N-1)^{1/2} r(x_1, \ldots, x_k)$. We employ the obvious notations for these values t_{MIN}, t_{MAX}, and C_{MIN}, C_{MAX}.

The usefulness of such calculations is immediately apparent. Suppose we want to test the alternative that Population 1 values are "larger" than for Population 0 using a scored t statistic and the t distribution as the approximating null distribution. If we find $t_{\mathrm{MIN}} > t_{N-2}^{\alpha}$, then we would know that *any possible* scoring system would produce a significant level-α result; that is, rejection of the null hypothesis does *not* depend on the choice of scores. Conversely, if $t_{\mathrm{MAX}} < t_{N-2}^{\alpha}$, then *no choice of scores* can produce a significant α-level result for this one-sided null hypothesis. However, if the optimized t statistics straddle t_{N-2}^{α} (i.e., $t_{\mathrm{MIN}} < t_{N-2}^{\alpha} < t_{\mathrm{MAX}}$), the choice of reasonable or meaningful scoring systems becomes crucial, and attention should be paid to the values $x_1, \ldots, x_k$ at which the minimum and maximum occur. Since the t distribution is an approximation to the true null distribution, care must be taken in this approach when exact (permutation) distributions are used to compute the null distribution of $t(x_1, \ldots, x_k)$. For example, conceptually both t_{MIN} and t_{MAX} could slightly exceed t_{N-2}^{α}, and yet the exact distribution for specific scores could yield a level of significance greater than α.

As noted in Section 2, we see that the t statistic and the one-sided C statistic are both monotonically increasing functions of $r(x_1, \ldots, x_k)$. Hence, if we compute

$$r_{\mathrm{MAX}} \equiv \max r(x_1, \ldots, x_k), \tag{3.1}$$

where the maximum is taken over the set $\{x_1 \leq \cdots \leq x_k, x_1 \neq x_k\}$, and

$$r_{\mathrm{MIN}} \equiv \min r(x_1, \ldots, x_k), \tag{3.2}$$

where the minimum is taken over the set $\{x_1 \leq \cdots \leq x_k, x_1 \neq x_k\}$, we can immediately use these to compute t_{MAX}, t_{MIN}, and C_{MAX}, C_{MIN}. For this reason, we only consider optimizing $r(x_1, \ldots, x_k)$.

Due to the location and scale invariance of correlation, the optimizing values $x_1 \leq \cdots \leq x_k$ $(x_1 \neq x_k)$ are not unique. Specifically, if $r(x_1^*, \ldots, x_k^*) = r_{\mathrm{MAX}}$, then $r(\alpha x_1^* + \beta, \ldots, \alpha x_k^* + \beta) = r_{\mathrm{MAX}}$ for any choices of $\alpha > 0$ and β. For ease of interpretation we usually require that the optimum scores satisfy $x_1 = 0$ and $x_k = 1$. At the same time, this requirement also ensures that the scores will be nondegenerate. (For convenience, in the proofs in the Appendix, we sometimes use other constraints for $x_1, \ldots, x_k$; for example, we employ the mean and variance constraints that $\Sigma(m_i + n_i)x_i/N = K_1$ and $\Sigma(m_i + n_i)x_i^2/N = K_2$, respectively.)

The notion of *stochastic ordering* plays an important role in whether or not the optimized correlations straddle 0. The data from Population 1 are said to be stochastically greater than the data from Population 0 if

$$(n_j + \cdots + n_k)/n \geq (m_j + \cdots + m_k)/m, \tag{3.3}$$

for $j = 2, \ldots, k$, that is, if the empirical distribution of the data from Population 1 is stochastically greater than the empirical distribution of the data from Population 0. If the inequality in (3.3) is reversed, then the Population 1 data are said to be stochastically smaller than the Population 0 data. If neither stochastically smaller or larger conditions hold, the data from these two populations are incomparable with respect to stochastic ordering. One well-known stochastic ordering result states that the data from Population 1 are stochastically greater (smaller) than that of Population 0 if and only if $r(x_1, \ldots, x_k) \geq 0$ (≤ 0), for all possible increasing scores.

Consider again the t_{MAX} discussion. If the Population 1 data are stochastically greater than the Population 0 data, then we know $t_{\mathrm{MIN}} \geq 0$. On the other hand, if the data from Population 1 are stochastically smaller than those from Population 0 then $t_{\mathrm{MAX}} \leq 0$, and if they are incomparable, $t_{\mathrm{MIN}} \leq 0 \leq t_{\mathrm{MAX}}$. This leads us to the important observation: *If the two data sets are stochastically incomparable, then there exist scores for which the t test will not reject the null hypothesis against a one-sided or two-sided alternative (for any $\alpha \leq .5$).*

Other solutions to similar types of optimization problems have been considered in differing, but related, contexts by Breiman and Friedman (1985), by Nishisato (1980), and by Kimeldorf, May, and Sampson (1982).

4. COMPUTATION

In this section, we describe in detail the steps required to compute the scores that minimize and maximize the correlation $r(x_1, \ldots, x_k)$. As will be illustrated, the computation of these scores can easily be done with a simple hand calculator. The proofs that these computations do give scores that maximize and minimize the correlations are left to the Appendix. We begin by checking to see whether the empirical distributions of the two populations are stochastically ordered, and if so, in what direction. This entails checking (3.3) with both directions for the inequalities. Note that this check depends only on the observed frequencies $m_1, \ldots, m_k$ and $n_1, \ldots, n_k$ and not on any particular choice of scores.

For clarity, we separately consider the computation in three cases: Population 0 data are (1) stochastically greater than, (2) stochastically less than, or (3) stochastically incomparable with Population 1 data. Consider now case (3) where the populations are incomparable. Then the scores $0 = x_1^* \leq \cdots \leq x_k^* = 1$ that maximize $r(x_1, \ldots, x_k)$ are found by first solving for the scores $y_1^* \leq \cdots \leq y_k^*$ that minimize the weighted sum of squares

$$\sum_{i=1}^{k} (m_i + n_i)\{n_i/(m_i + n_i) - y_i\}^2, \tag{4.1}$$

among $y_1 \leq \cdots \leq y_k$. These scores $y_1^* \leq \cdots \leq y_k^*$ are the isotonic regression of $n_i/(m_i + n_i)$ with weights $m_i + n_i$. There are a variety of algorithms (see Robertson, Wright, and Dykstra 1988) that can be used to compute $y_1^*, \ldots, y_k^*$. Among these is the simple and elegant pool adjacent violators algorithm (PAVA) that we use in our examples. The fact that the populations are not stochastically ordered insures that $y_k^* > y_1^*$; thus, a simple linear transformation

$$x_i^* = (y_i^* - y_1^*)/(y_k^* - y_1^*), \qquad i = 1, \ldots, k, \tag{4.2}$$

gives the scores, with $x_1^* = 0 \leq \cdots \leq x_k^* = 1$, where $r(x_1^*, \ldots, x_k^*) = r(y_1^*, \ldots, y_k^*)$. The scores $0 = z_1^* \leq \cdots \leq z_k^* = 1$ that minimize $r(x_1, \ldots, x_k)$ are found similarly by first finding scores $y_1^* \le \cdots \le y_k^*$ that minimize the weighted sum of squares

$$\sum_{i=1}^{k} (m_i + n_i)\{m_i/(m_i + n_i) - y_i\}^2, \qquad (4.3)$$

among $y_1 \le \cdots \le y_k$. Here, $y_1^*, \ldots, y_k^*$ are the isotonic regression of $m_i/(m_i + n_i)$ with weights $m_i + n_i$ and can also be computed using PAVA. Again, the absence of stochastic ordering ensures that $y_k^* - y_1^* > 0$, and the linear transformation of (4.2) gives the required scores $0 = z_1^* \le \cdots \le z_k^* = 1$.

Now consider the case that the Population 1 data are stochastically greater than the Population 0 data. The scores $0 = x_1^* \le \cdots \le x_k^* = 1$ that maximize the correlation are found using the procedures in (4.1) and (4.2). The scores $0 = z_1^* \le \cdots \le z_k^* = 1$ that minimize the correlation cannot be found using isotonic regression techniques. These scores are shown, however, to occur at a *monotone extreme point,* that is, a point where $z_i^* = 0$ for $1 \le i \le j$ and $z_i^* = 1$ for $j + 1 \le i \le k$ for some $j = 1, \ldots, k - 1$. Therefore, simple calculation of the correlation for the $k - 1$ monotone extreme points will suffice to identify $z_1^*, \ldots, z_k^*$, the monotone extreme point with the smallest correlation. Note that the monotone extreme points correspond to all ordinal dichotomizations of the levels.

Finally, consider the case that the Population 0 data are stochastically greater than Population 1 data. Scores $0 = x_1^* \le \cdots \le x_k^* = 1$, that maximize the correlation, now are a monotone extreme point. Thus one needs to compute $r(x_1, \ldots, x_k)$ for the $k - 1$ monotone extreme points and let $x_1^*, \ldots, x_k^*$ be the monotone extreme point with the largest correlation. The scores $0 = z_1^* \le \cdots \le z_k^* = 1$ that minimize correlation are found using the isotonic regression techniques given in (4.3).

We conclude this section with three examples, which for explicitness we analyze using t tests. The first set of data (Table 2) is based on data from Koopmans (1987, p. 425) that were collected to compare the severity of case dispositions in the juvenile courts between boys and girls in a certain county in New Mexico. A random sample of court records of 152 males and 156 females was selected and classified as "(1) counseled and released; (2) one intervention by the probation department; (3) two or more interventions; and (4) referral to juvenile court."

For these data the empirical distribution for males is stochastically larger than the empirical distribution for females. The scores $x_1^* = x_2^* = x_3^* = 0$, $x_4^* = 1$ maximize the correlation ($r_{\text{MAX}} = -.235$) and can be found by calculating the correlation for each monotone extreme point. To compute scores to minimize the correlation, we find the isotonic regression of $m_i/(m_i + n_i)$ with weights $m_i + n_i$. Since the scores $m_i/(m_i + n_i)$ are increasing in i (see Table 2), the isotonic regression is $y_i^* = m_i/(n_i + m_i)$, $i = 1, \ldots, 4$. Therefore, $z_1^* = 0$, $z_2^* = .3836$, $z_3^* = .7937$, $z_4^* = 1.0000$, and $r_{\text{MIN}} = -.317$. Both $t_{\text{MIN}} = -5.85$ and $t_{\text{MAX}} = -4.22$ indicate that the null hypothesis that boys and girls receive the same court treatment should be rejected at any reasonable level of significance.

The second set of data, using the data (Devore and Peck 1986, p. 636) given in Table 3, compares smokers' and nonsmokers' opinions of an antismoking ad. The empirical distributions of opinion for smokers and nonsmokers are not stochastically ordered. The isotonic regression $y_1^*, \ldots, y_k^*$ partitions $\{1, \ldots, k\}$ into blocks called level sets. On each level set the value of y is the weighted average of $n_i/(m_i + n_i)$ over the level set using the weights $(m_i + n_i)$. PAVA identifies the level sets in a sequence of steps. At the first step look for a pair where $n_i/(m_i + n_i) > n_{i+1}/(m_{i+1} + n_{i+1})$ [or $m_i/(m_i + n_i) > m_{i+1}/(m_{i+1} + n_{i+1})$]. These two values are pooled to form one block with weights $(m_i + n_i + m_{i+1} + n_{i+1})$ and value $(n_i + n_{i+1})/(m_i + n_i + m_{i+1} + n_{i+1})$ [or $(m_i + m_{i+1})/(m_i + n_i + m_{i+1} + n_{i+1})$]. At each successive step there is one fewer value, and a new pair is identified to pool. The algorithm stops when values for the blocks are nondecreasing. From Table 4 we see that the isotonic regression of $n_i/(m_i + n_i)$ with weights $(m_i + n_i)$ is $y_1^* = y_2^* = y_3^* = .7260$, $y_4^* = .7439$, and $y_5^* = .7841$. Thus $x_1^* = x_2^* = x_3^* = 0$, $x_4^* = .3086$, and $x_5^* = 1$ maximize the correlation, and $r_{\text{MAX}} = .054$ with $t_{\text{MAX}} = 1.045$. The isotonic regression of $m_i/(n_i + m_i)$ with weights $(m_i + n_i)$ is $y_1^* = .2051$, $y_2^* = .2500$, $y_3^* = y_4^* = y_5^* = .2698$, and, therefore, the scores $z_1^* = 0$, $z_2^* = .7492$, and $z_3^* = z_4^* = z_5^* = 1$ minimize the correlation with $r_{\text{MIN}} = -.042$ and $t_{\text{MIN}} = -.811$. Neither t_{MIN} nor t_{MAX} exceed any usual critical value ($\alpha < .1$), thereby showing that there is no difference between smokers' and nonsmokers' responses to the ad.

We note that, in this example, because the data were not stochastically comparable, we knew a priori that there were scores that would not allow rejecting a two-sided or one-sided alternative. As the calculations indicate, however, no possible monotonically increasing scoring can produce a significant result.

The third set of data (Agresti 1984, p. 30) compares changes in size of ulcer crater under two treatments A and B; see Table 5.

For this data the empirical distribution of crater size under treatment A can be shown to be stochastically less than the empirical distribution of crater size under treatment B.

Table 2. Severity of Juvenile Court Case Dispositions

	Levels			
	1	2	3	4
Males (0)	63	41	18	30
Females (1)	107	35	7	7
$m_i/(n_i + m_i)$	.3706	.5395	.7200	.8108

Source: Koopmans (1987).

Table 3. Opinions of Antismoking Ad

	Opinions				
	Strongly dislike	Dislike	Neutral	Like	Strongly like
Smoker (0)	8	14	35	21	19
Nonsmoker (1)	31	42	78	61	69

Source: Devore and Peck (1986)

Table 4. Pool Adjacent Violators Algorithm for Data in Table 3

i	1	2	3	4	5
$n_i/(m_i + n_i)$	.7949	.7500	.6903	.7439	.7841
First pool	$(n_1 + n_2)/(m_1 + n_1 + m_2 + n_2) = 73/95 = .7684$				
Second pool	$(n_1 + n_2 + n_3)/(m_1 + n_1 + m_2 + n_2 + m_3 + n_3) = 151/208 = .7260$				
$m_i/(m_i + n_i)$	.2051	.2500	.3097	.2561	.2159
First pool			$(m_3 + m_4)/(m_3 + n_3 + m_4 + n_4) = 56/195 = .2872$		
Second pool			$(m_3 + m_4 + m_5)/(m_3 + n_3 + m_4 + n_4 + m_5 + n_5) = 75/278 = .2698$		

The scores $z_1^* = z_2^* = z_3^* = 0$, and $z_4^* = 1$ that minimize the correlation ($r_{\text{MIN}} = .177$) can again be found by calculating the correlations for all monotone extreme points. To maximize the correlation, we find the isotonic regression of $n_i/(m_i + n_i)$ with weights $m_i + n_i$, to obtain $x_1^* = 0$, $x_2^* = .4164$, $x_3^* = x_4^* = 1$, and $r_{\text{MAX}} = .304$. Here $t_{\text{MIN}} = 1.42$ and $t_{\text{MAX}} = 2.508$. Assuming a two-sided alternative with a traditional level of significance, we find that there are, in this straddling case, some scores that produce significance and some scores that do not produce significance.

5. DISCUSSION

Rather than maximize or minimize the t statistic or C statistic over the set of *all* possible assignments of increasing scores, the set of possible assignments might be constrained further. For example, we may want to restrict our attention to "symmetric" scorings in which $x_{i+1} - x_i = x_{k-i+1} - x_{k-i}$. Alternatively, we might want to restrict our attention to scorings satisfying certain inequality constraints, such as $x_3 - x_2 \leq 2(x_2 - x_1)$. If the set of scorings is restricted, then the simple algorithms presented here are no longer applicable. On the other hand, as long as the constrained set $x_1, x_2, \ldots, x_k$ of scores form a convex subspace of k space, the values r_{MIN} and r_{MAX}, as well as the scorings at which they are obtained, can be computed. An algorithm useful in this case is presented in Robertson et al. (1988), and further algorithms allowing the computation of both r_{MIN} and r_{MAX} appear in a dissertation by Gautam (1991).

In our experience with applying our proposed min and max technique to reported data in the literature, we have found that the nonstraddling cases occur more often than we might have originally expected. In such cases we believe that our approach allows for a strong scientific statement concerning the significance or insignificance of the difference between the two populations or treatments.

Obviously, the straddling case is more problematic in its interpretation for the statistician doing the analysis. To be clear, we are not in any way advocating that the optimizing scores be used to suggest interpretable scores that would make results significant or insignificant according to the experimental purpose. We assert, however, that the scientific meaning of the scores that produce both significance and insignificance must be examined in the context of the experiment for their relevance. More experience with our approach may indicate further avenues of interpretation in this ambiguous case. Perhaps one might use the degree of overlap of the max and min statistics relative to their appropriate critical value as an indication of the strength of experimental evidence. This overlap could be measured on the scale of the statistic or on the scale of "p values." Another, perhaps more interesting, approach is rooted in the ideas of Diaconis and Efron (1985). One might compute the $(k - 2)$-dimensional volume of the increasing scores that produce significance and compare it in a suitable fashion to the volume of increasing scores that do not produce significance. Nonetheless, this straddling case, we expect, will always be somewhat difficult in its interpretability due to the inherent ambiguity in the data.

Table 5. Change in Size of Ulcer Crater

Treatment	Larger	<2/3 Healed	≥2/3 Healed	Healed
A	12	10	4	6
B	5	8	8	11

Source: Agresti (1984).

In summary, we note that if, for example, both t_{MAX} and t_{MIN} fall (or don't fall) in the usual t critical region, we know that any increasing scoring must produce a significant (or insignificant) result based on a t test with that scoring. Consequently, our systematic approach is preferable to one where the statistician nonsystematically tries a number of possible different scorings. Thus we think that any analyses of two-sample ordinal data should include the maximum and minimum values of the appropriate test statistic, as well as the increasing scores that produce each of these. Further development is required if we attempt to use optimized statistics for testing. If, for example, we use t_{MAX} and t_{MIN} as test statistics in their own right, their asymptotic distributions under the null hypothesis are obviously not a standard t distribution. (For fixed sample size, however, under the null hypothesis, we note that the distribution of t_{MIN} is stochastically less than the distribution of the t statistic with any scores which, in turn, is stochastically less than the distribution of t_{MAX}). Finally, we caution that, in the straddling case, care must be taken to scientifically justify the choice of scores on which inference is based.

APPENDIX: PROOFS

Let (U, V) have joint probability mass function based on the observed frequencies and define $p_{0j} = P(U = 0, V = j) = m_j/N$

and $p_{1j} = P(U = 1, V = j) = n_j/N$, for $1 \leq j \leq k$. For $i = 0, 1$, let V_i be the random variables with $P(V_i = j) = P(V = j \mid U = i)$. Finding the scores that maximize and minimize the correlation between populations is equivalent to finding functions ϕ^* and ϕ_*, respectively, to maximize and minimize corr(U, $g(V)$) among the scoring functions $g \in C$, where $C = \{g:\{1, \ldots, k\} \to \mathbf{R}^1$, g is nondecreasing, $g(1) < g(k)\}$. Stochastic ordering (or lack thereof) of V_1 and V_0 is equivalent to stochastic ordering of the empirical distributions based on the data from the two populations. We note that minimizing corr(U, $g(V)$) is equivalent to maximizing corr($1 - U$, $g(V)$), that is, ϕ_* can be found by relabeling the populations and then maximizing the correlation based on data from the relabeled populations. Thus we concentrate on solving for ϕ^* that maximizes corr(U, $g(V)$) among $g \in C$.

We consider two cases when solving for ϕ^*: V_1 is not stochastically less than V_0 (i.e., either V_1 is stochastically greater than, but not equal in distribution to V_0, or V_1 and V_0 are not stochastically ordered), and the case V_1 is stochastically less than V_0. Solutions for ϕ_* follow as corollaries for the analogous cases V_1 is not stochastically greater than V_0, and V_1 is stochastically greater than V_0.

Theorem A.1. Suppose V_1 is not stochastically less than V_0. Then ϕ^* is the isotonic regression of $n_i/(m_i + n_i)$ with weights $m_i + n_i$.

Proof. Following the arguments of Robertson et al. (1988, p. 379), we show that maximizing corr(U, $g(V)$) among $g \in C$ is equivalent to maximizing corr(U, $g(V)$) among $g \in C'$, where $C' = \{g \in C: E[g(V)] = E[U] \text{ and } E[Ug(V)] = E[g^2(V)]\}$. To see that these two problems are equivalent, let $h \in C$. If cov(U, $h(V)$) ≤ 0, then h is not a candidate to maximize corr(U, $h(V)$) because V_1 not stochastically less than V_0 implies the existence of a $g \in C$ such that cov(U, $g(V)$) > 0. If cov(U, $h(V)$) > 0 then define

$$g(V) = E(U) + (h(V) - E[h(V)])\mathrm{cov}(U, h(V))/\mathrm{var}(h(V)).$$

Then it is clear that $g \in C'$ and corr(U, $g(V)$) = corr(U, $h(V)$). Let $g \in C'$

$$\begin{aligned}\mathrm{corr}(U, g(V)) &= \mathrm{cov}(U, g(V))/[\mathrm{var}(U)\ \mathrm{var}(g(V))]^{1/2} \\ &= [\mathrm{cov}(U, g(V))/\mathrm{var}(U)]^{1/2} \\ &= [1 - E(U - g(V))^2/\mathrm{var}(U)]^{1/2}. \qquad \text{(A.1)}\end{aligned}$$

In addition

$$\begin{aligned}E[(U - g(V))^2] &= \sum_{i=1}^{k} [g^2(i)p_{0i} + (1 - g(i))^2 p_{1i}] \\ &= \sum_{i=1}^{k} \{(p_{0i} + p_{1i})[p_{1i}(p_{1i} + p_{0i})^{-1} - g(i)]^2 \\ &\quad + p_{1i}p_{0i}[p_{1i} + p_{0i}]^{-1}\}, \qquad \text{(A.2)}\end{aligned}$$

where k is a constant not depending on g. Thus, from (A.1) and (A.2), we see that maximizing corr(U, $g(V)$) is equivalent to minimizing

$$\begin{aligned}&\sum_{i=1}^{k} \{(p_{0i} + p_{1i})p_{1i}/(p_{1i} + p_{0i}) - g(i)\}^2 \\ &\qquad = N^{-1} \sum_{i=1}^{k} (m_i + n_i)\{n_i/(m_i + n_i) - g(i)\}^2 \qquad \text{(A.3)}\end{aligned}$$

among $g \in C'$. Let ϕ^* be the isotonic regression of $n_i/(m_i + n_i)$ with weights $(m_i + n_i)$. Simple calculations and the properties of ϕ^* summarized in Robertson et al. (1988, theorems 1.3.2 and 1.3.3) verify that ϕ^* does minimize (A.3) among $g \in C'$. Thus

$$\mathrm{corr}(U, \phi^*(V)) = \max_{g \in C} \mathrm{corr}(U, g(V)).$$

Corollary A.1. Suppose V_1 is not stochastically greater than V_0, then ϕ_* is the isotonic regression of $m_i/(m_i + n_i)$ with weights $m_i + n_i$.

Thus Theorem A.1 and Corollary A.1 provide solutions for both ϕ_* and ϕ^* when V_1 and V_0 are not stochastically ordered.

Now suppose that V_1 is stochastically less than V_0. Corollary A.1 gives ϕ_*; however, the previous isotonic regression techniques cannot be used to find ϕ^*. To find ϕ^*, we need the following definition and theorem.

Definition. Let f: $S \to \mathbf{R}$, where S is a nonempty convex subset of $\mathbf{R}^k$. The function f is said to be quasiconvex if, for each $\mathbf{x}$, $\mathbf{y} \in S$,

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \max(f(\mathbf{x}), f(\mathbf{y})), \qquad \lambda \in [0, 1].$$

Theorem A.2. (Bazaraa and Shetty 1979, Theorem 3.5.3) Suppose f: $S \to \mathbf{R}$ is a quasiconvex function and $S \subseteq \mathbf{R}^k$ is a compact polyhedral. Then f attains its maximum at an extreme point of S.

Because correlation is scale and location invariant, maximizing corr(U, $g(V)$) among $f \in C$ is equivalent to maximizing corr(U, $g(V)$) among $g \in \mathbf{H} \equiv \{h: h \in C, h(1) = 0 \text{ and } h(k) = 1\}$.

Let $\mathbf{g} = (g(1), \ldots, g(k))$ be the vector representation of $g \in C$ and define $S = \mathbf{H}$ and $f(\mathbf{g})$ = corr(U, $g(V)$). We additionally use the notation cov(U, $\mathbf{g}$), var($\mathbf{g}$) to denote the obvious quantities. Then S is the compact polyhedral in $\mathbf{R}^k$ defined by the linear constraints $g(1) = 0$, $g(k) = 1$, and $g(i) - g(i - 1) \geq 0$, for $2 \leq i \leq k$. The extreme points of S are $\{\mathbf{g}: g(i) = 0, 1 \leq i \leq j, g(i) = 1, j + 1 \leq i \leq k, \text{ for } 1 \leq j \leq k - 1\}$. To see that $f(\mathbf{g})$ is quasiconvex, note that, because V_1 is supposed stochastically less than V_0, cov(U, $\mathbf{g}$) ≤ 0, for all $\mathbf{g} \in S$, and cov(U, $\mathbf{g}$) is linear in $\mathbf{g}$. Furthermore, the denominator $\sqrt{\mathrm{var}(U)\mathrm{var}(\mathbf{g})}$ of $f(\mathbf{g})$ is strictly positive and easily shown to be convex for $\mathbf{g} \in S$. Since $f(\mathbf{g})$ is the ratio of a negative concave function of $\mathbf{g}$ and a strictly positive convex function of $\mathbf{g}$, then $f(\mathbf{g})$ is quasiconvex (Bazaraa and Shetty 1979, problem 3.39). Thus $f(\mathbf{g})$ attains its maximum among the extreme points of S, and we have the following result.

Theorem A.3. Suppose V_1 is stochastically less than V_0. Then for some $1 \leq j \leq k - 1$, $\phi^*(i) = 0$, for $1 \leq i \leq j$, and $\phi^*(i) = 1$, for $j + 1 \leq i \leq k$.

Corollary A.3. Suppose V_1 is stochastically greater than V_0. Then for some $i \leq j \leq k - 1$, $\phi_*(i) = 0$ for $1 \leq i \leq j$ and $\phi_*(i) = 1$ for $j + 1 \leq i \leq k$.

Finally, Theorem A.1 and Corollary A.3 give solutions for ϕ_* and ϕ^* when V_1 is stochastically greater than V_0. Similarly, the solutions for ϕ_* and ϕ^* are given by Theorem A.3 and Corollary A.1 when V_1 is stochastically less than V_0.

[*Received February 1990. Revised May 1991.*]

REFERENCES

Agresti, A. (1984), *Analysis of Ordinal Categorical Data,* New York: John Wiley.

——— (1990), *Categorical Data Analysis,* New York: John Wiley.

Agresti, A., Mehta, C. R., and Patel, N. R. (1990), "Exact Inference for Contingency Tables With Ordered Categories," *Journal of the American Statistical Association,* 85, 453–458.

Bazaraa, M. S., and Shetty, C. M. (1979), *Nonlinear Programming, Theory and Applications,* New York: John Wiley.

Breiman, L., and Friedman, J. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association,* 80, 580–598.

Conover, W. J., and Iman, R. L. (1981), "Rank transformations as a

Bridge Between Parametric and Nonparametric Statistics," *The American Statistician*, 35, 124–129.

Devore, J., and Peck, R. (1986), "*Statistics, the Exploration and Analysis of Data*," St. Paul, MN: West.

Diaconis, P., and Efron, B. (1985), "Testing for Independence in a Two-Way Table: New Interpretations of the Chi-Square Statistics," *The Annals of Statistics*, 13, 845–874.

Fleiss, J. L. (1981), *Statistical Methods for Rates and Proportions* (2nd ed.), New York: John Wiley.

Gautam, S. (1991), "Applications of the t, T^2 and F Statistics to Ordinal Categorical Data," unpublished Ph.D. dissertation, The University of Texas at Dallas, Program in Mathematical Sciences.

Graubard, B. I., and Korn, E. L. (1987), "Choice of Column Scores for Testing Independence in Ordered $2 \times k$ Contingency Tables," *Biometrics*, 43, 471–476.

Halpern, J. (1982), "Maximally Selected Chi-Square Statistics for Small Samples," *Biometrics*, 38, 1017–1023.

Hettmansperger, T. P. (1984), *Statistical Inference Based Upon Ranks*, New York: John Wiley.

Kimeldorf, G., May, J., and Sampson, A. R. (1982), "Concordant and Discordant Monotone Correlations and Their Evaluation Using Nonlinear Optimization," in *TIMS Studies in the Management Sciences. Optimization in Statistics* (vol. 19), eds. S. H. Zanakis and J. S. Rustagi, New York: North Holland, 117–130.

Koopmans, L. H. (1987), *Introduction to Contemporary Statistical Methods* (2nd ed.), Boston: Duxbury.

Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

Mantel, N. (1963), "Chi-Squared Tests With One Degree of Freedom: Extensions of the Mantel–Haenszel Procedure," *Journal of the American Statistical Association*, 58, 690–700.

Miller, R. G. (1986), *Beyond ANOVA: Basics of Applied Statistics*, New York: John Wiley.

Miller, R. G., and Siegmund, D. (1982), "Maximally Selected Chi Square Statistics," *Biometrics*, 38, 1011–1016.

Nishisato, S. (1980), *Analysis of Categorical Data: Dual Scaling and Its Applications*, Toronto: University of Toronto Press.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, New York: John Wiley.

SAS Institute, Inc. (1985), *SAS User's Guide: Statistics Version 5 Edition*, Cary, NC, Author.

Woodward, J., and Overall, J. (1977), "The Significance of Treatment Effects in Ordered Categorical Data," *Journal of Psychiatric Research*, 13, 169–177.