Journal of Statistical Planning and
Inference 107 (2002) 219–236

journal of
statistical planning
and inference

www.elsevier.com/locate/jspi

Min and max scorings for two sample partially ordered categorical data

Allan R. Sampson[a,*,1], Harshinder Singh[b,c]

[a]Department of Statistics, University of Pittsburgh, CL 2700, Pittsburgh, PA 15260, USA
[b]Department of Statistics, West Virginia University, Morgantown, WV 26506, USA
[c]Biostatistics Branch, Health Effects Laboratory Division, NIOSH, Morgantown, WV 26505, USA

Abstract

For certain types of designed experiments the outcome variables are ordinal categorical variables. Such experiments are typical in many behavioral science and social science settings, as well as in certain types of clinical trials. The representation, analysis and summarization of the results of such experiments can be difficult and complex. We present one analytical approach to these settings that uses scalings for a simpler presentation of the results. The resulting scalings can be viewed as a technique for reducing the complexity of the multivariate ordinal response data and providing insightful data summaries.

Utilizing some of the results for designed experiments with univariate ordinal responses, we introduce new techniques to handle responses taking values in a partially ordered set. These latter results are applied to scaling multivariate ordinal data. Several data sets are analyzed using this new approach. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Partial order; Ordinal data; Scoring; t-statistic; Cochran–Armitage statistic; Stochastic ordering; Chi-square; Max-correlation

---

1. Introduction

Consider data samples from two populations where each observation falls in one of k distinct categories. A typical goal in this setting is to assess whether or not there is a difference between the two populations and to describe the difference if it is thought to exist. In the case where there is no ordering among the k categories, i.e., all the categories are nominal, the usual chi-square test is most commonly used.

If the categories are completely ordered, i.e., ordinal data, then among the commonly used tests are the Wilcoxon–Mann–Whitney test, the Cochran–Armitage procedure, the

---

[*] Corresponding author. Tel.: +1-412-624-8372; fax: +1-412-648-8814.
E-mail address: asampson@stat.pitt.edu (A.R. Sampson).

[1] Research of A.R. Sampson is supported by National Science Foundation Grant DMS-0072207.

PII: S0378-3758(02)00254-9

scored t-test, and tests based on certain log-linear models with scores assigned to the categories. (see Agresti, 1984, 1990) To handle the ambiguities arising from the choice of testing procedures and scorings, Kimeldorf et al. (1992) (henceforth, KSW, 1992) presented a new approach to summarizing the comparison of two populations. They established that many of these commonly used test statistics are essentially monotone functions of a "key statistic" (which is a certain correlation) with increasing scores assigned to the ordered levels of the categories. Moreover, they noted that the scored t-test with various scorings basically includes these commonly employed tests. Using techniques of ordered restricted statistical inference, KSW (1992) obtained the minimum and maximum of the key statistic over all possible assignments of nondegenerate increasing scores and the values at which these optima occurred. This led them to propose reporting both the maximum and the minimum value of the corresponding scored t-statistic (and sometimes the scored Cochran–Armitage test statistic). If the range of min and max value does not include the critical value of the test statistic (they term this case "nonstradling"), then it can be immediately concluded that the result of the analysis remains the same no matter the choice among these tests and the increasing scores used. However, if the range includes the critical value (termed the "straddling" case), the choice of procedures and scores used in the analysis must be carefully justified. They demonstrated their technique through a number of examples.

In this paper we consider a more general version of the KSW (1992) approach which allows us to handle more complicated categorical data in the two-sample setting. Two situations of interest arise in practice and are somewhat difficult to analyze. One is where we observe multiple and different ordinal variables for each experimental unit. The other is where we observe a single more complex categorical variable where the levels are tree-ordered. An example of the former occurs when we wish to compare two clinical treatments based upon observing for each patient two ordinal variables, namely, the physician's global improvement rating and the patient's global improvement rating where each evaluation is on the scale: excellent, very good, good, fair, and poor. An example of a tree ordering is a categorization such as no anxiety, moderate anxiety, moderate anxiety with depression, and severe anxiety.

Both these examples fall into the general setting of a partial order on a set of k categories. This leads us to examining issues from KSW (1992) in the context of a partial order on the k categorical variables.

A natural way to tackle this problem is to assign justifiable scores to levels of all the categories according to the partial order on the levels and then analyze the data using one of the previously noted procedures. Since scoring of the partially ordered levels is more problematic than for the case of completely ordered levels, it is again important to find the minimum and maximum of the two-sample test statistic over all possible assignments of scores to levels determined by the partial order on the categories. If it turns out to be the nonstradling case, it can be once again concluded that the result of the analysis does not depend on the choice of scores assigned. For the straddling case, the choice of scores in the analysis needs to be carefully justified.

Assuming the partial order, we give a technique to find the min and max of standard two-sample test statistics over all possible nondegenerate scores allowed by the partial order on the levels of the categories. In Section 2, we give the max and min value of

Table 1
Data format

		Levels				
		L_1	L_2	…	L_k	Total
Populations	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

the "key statistic" over all possible scores assigned to the levels allowed by the partial order. In Section 3, we demonstrate our procedure on some examples where levels of the categories are partially ordered.

These results generalize KSW (1992) in that ordinal data seen as a complete ordering is a particular case of a partial order.

2. Min and max of correlation for partially ordered categorical levels

Following the same notation as KSW (1992), we suppose the data are drawn from two populations denoted for convenience as 0 and 1, where each observation falls into one of the k categories. Denote the levels of the k categories by $L_1, L_2, \ldots, L_k$. Let $\preccurlyeq$ denote the underlying experimental ordering, where we assume that $\preccurlyeq$ is a partial order, i.e., $\preccurlyeq$ is reflexive, transitive and antisymmetric, but there may be noncomparable levels.

We represent such two-sample data schematically in Table 1.

Proceeding in a natural way, we let scores $x_1, x_2, \ldots, x_k$ be assigned to the levels $L_1, L_2, \ldots, L_k$ as allowed by the partial order $\preccurlyeq$ on $L_1, L_2, \ldots, L_k$ making sure that not all x_i's are equal. (For example, if $k = 3$, $L_1 \preccurlyeq L_2$, $L_1 \preccurlyeq L_3$, and L_2, L_3 are not comparable, then x_1, x_2, x_3 satisfy only $x_1 \leqslant x_2$, $x_1 \leqslant x_3$ and that not all x_1, x_2, x_3 are equal.) The t-statistic computed with scores $x_1, x_2, \ldots, x_k$ assigned to the levels $L_1, L_2, \ldots, L_k$ is given by

$$t(x_1, x_2, \ldots, x_k) = \sqrt{m+n-2}\,\frac{r(x_1, x_2, \ldots, x_k)}{\sqrt{1 - r^2(x_1, x_2, \ldots, x_k)}}, \tag{2.1}$$

where $r = r(x_1, x_2, \ldots, x_k)$ is the Pearson coefficient of correlation based on the scores $x_1, x_2, \ldots, x_k$ and values 0 and 1 assigned to the populations and is given by

$$r = \sqrt{\frac{mn}{m+n}}\,\frac{A_1 - A_0}{\sqrt{\sum (m_i + n_i) x_i^2 - \frac{1}{m+n}(mA_0 + nA_1)^2}}, \tag{2.2}$$

where $A_0 = (1/m)\sum m_i x_i$ and $A_1 = (1/n)\sum n_i x_i$ (see KSW, 1992). Notationally if the scores x_i are given by a function $f(L_i)$, then we sometimes write $t(f)$ instead of $t(x_1, \ldots, x_n)$. Because the data are not normal, the sampling distribution of the t-statistic is not the t-distribution (except for large samples).

Similarly, the (two-sided) Cochran–Armitage statistic $C(x_1,x_2,\dots,x_k)$ is given by

$$C(x_1,x_2,\dots,x_k)=(m+n-1)r^2. \tag{2.3}$$

As noted by KSW (1992), t and the one-sided version of C are both monotonically increasing functions of $r(x_1,x_2,\dots,x_k)$, so that we need only give a procedure to find $r_{\max}$ and $r_{\min}$ over all nondegenerate scores $x_1,x_2,\dots,x_k$ allowed by the partial order $\preccurlyeq$. These can then be straightforwardly used to compute $t_{\max}$, $t_{\min}$ and $C_{\max}$, $C_{\min}$. In view of the location and scale invariance of r, we assume that the optimized scores satisfy $\min(x_1,x_2,\dots,x_k)=0$ and $\max(x_1,x_2,\dots,x_k)=1$. This also ensures that scores are nondegenerate.

Let $X=\{L_1,L_2,\dots,L_k\}$ denote the set of levels of the k categories. The following definitions are needed in the sequel.

Definition 2.1. A subset L of X is called a lower set with respect to the partial order $\preccurlyeq$ if $L_i\in X$, $L_j\in L$ and $L_i\preccurlyeq L_j$ imply $L_i\in L$. A subset U of X is called an upper set if $L_i\in U$ and $L_j\in X$ and $L_i\preccurlyeq L_j$ imply $L_j\in U$.

X is a trivial upper set as well as trivial lower set, as is ϕ. An upper (lower) set other than X and ϕ shall be called a nontrivial upper(lower) set.

We denote the class of all lower sets by $\mathscr{L}$ and the class of all upper sets by $\mathscr{U}$.

Upper and lower sets are important for defining stochastic ordering among data sets; furthermore, intersections of upper and lower sets—the so-called level sets—play a fundamental role in isotonic regression algorithms.

For example, for $k=3$, if $L_1\preccurlyeq L_2$, $L_1\preccurlyeq L_3$, and L_2, L_3 are not comparable with respect to the partial order, then

$$\mathscr{U}=\{\phi,\{L_2\},\{L_3\},\{L_2,L_3\},\{L_1,L_2,L_3\}\}$$

and

$$\mathscr{L}=\{\phi,\{L_1\},\{L_1,L_2\},\{L_1,L_3\},\{L_1,L_2,L_3\}\}.$$

For a given partial order, the enumeration and specification of all upper sets (or equivalently lower sets) may require careful algorithms and substantial computational time. For the matrix partial order, Sampson and Whitaker (1988) present some techniques that aid in counting and specifying upper sets. An indication of the amount of computation is that for the matrix order on a $4\times4\times4$ lattice, there are 232, 848 upper sets.

Definition 2.2. The data from population 1 are said to be stochastically larger with respect to partial order $\preccurlyeq$ than data from population 0, if

$$\frac{1}{n}\sum_{\{j:L_j\in U\}} n_j \geqslant \frac{1}{m}\sum_{\{j:L_j\in U\}} m_j \tag{2.4}$$

for every upper set $U\in\mathscr{U}$.

This notion of stochastic ordering is intimately related to stochastic ordering for random vectors (see Marshall and Olkin, 1979) where $\preccurlyeq$ is the usual coordinate-wise partial ordering. In Definition 2.2, if $\preccurlyeq$ is a complete ordering, the class of all upper sets is $\mathscr{U} = \{U_1, U_2, \ldots, U_k\}$ where $U_i = \{L_i, L_{i+1}, \ldots, L_k\}$ and the stochastic ordering with respect to $\preccurlyeq$ reduces to usual stochastic ordering between two data sets as utilized by KSW (1992).

Suppose we consider any consistent dichotomization of the data, by which we mean all levels L_i in some upper set U are considered as "success" and those in U^c are 'failure". As Sampson and Whitaker (1989) point out, the stochastic ordering of the data with population 1 being larger than population 0 guarantees that the observed probability of success for any reasonable dichotomization is higher for population 1 than is for population 0.

We use the terminology that a set of scores $x_1, \ldots, x_k$ preserves a partial order $\preccurlyeq$, or is consistent with the partial order $\preccurlyeq$, if whenever $L_i \preccurlyeq L_j$ it follows that $x_i \leqslant x_j$. These scores are assumed nontrivial or nondegenerate in the sense that not all scores are equal.

We have the following equivalence theorem for the stochastic ordering between data of population 1 and population 0 with respect to $\preccurlyeq$ and the nonnegativity of the key statistic $r(x_1, x_2, \ldots, x_k)$, the correlation coefficient. The proof is given in the Appendix A.

Theorem 2.1. *Data from population* 1 *are stochastically larger* (*smaller*) *with respect to* $\preccurlyeq$ *than data from population* 0, *if and only if* $r(x_1, x_2, \ldots, x_k) \geqslant (\leqslant) 0$ *for all possible scores* $x_1, x_2, \ldots, x_k$ *consistent with the partial order* $\preccurlyeq$.

The arguments of KSW (1992) show that, if the data of population 1 are stochastically larger than the data of population 0 with respect to the partial order $\preccurlyeq$, then $t_{\min} \geqslant 0$. On the other hand, if the population 0 data are stochastically larger with respect to the partial order than the population 1 data, then $t_{\max} \leqslant 0$. And if they are incomparable, then $t_{\min} \leqslant 0 \leqslant t_{\max}$, and thus for this case there exist scores consistent with the partial order for which the t-test will not reject the null hypothesis against a one- or two-sided alternative for any $\alpha \leqslant 0.5$.

Computations of $r_{\min}$ and $r_{\max}$ is different for the following three cases. The proofs that these computations do give scores that maximize and minimize $r(x_1, x_2, \ldots, x_k)$ are stated as theorems in Appendix A.

Case 1: *Population* 1 *data and population* 0 *data are stochastically incomparable with respect to the partial order* $\preccurlyeq$. For this case the first step is to obtain the scores $y_1^*, y_2^*, \ldots, y_k^*$ that minimize the expression

$$\sum_{i=1}^{k} \left(\frac{n_i}{m_i + n_i} - y_i \right)^2 (m_i + n_i) \tag{2.5}$$

among all scores y_i consistent with the partial order $\preccurlyeq$. These y_i^*'s are the generalized isotonic regression of $n_i/(m_i + n_i)$ with weights $(m_i + n_i)$ with respect to the partial order $\preccurlyeq$. The Minimum Lower Sets Algorithm (see Barlow et al., 1972 or Robertson et al., 1988) or the IBCR algorithm given by Block et al. (1994) can be used to compute the y_i^*'s, $i = 1, 2, \ldots, k$.

Then the linear transformation

$$x_i^* = \frac{y_i^* - \min(y_1^*, y_2^*, \ldots, y_k^*)}{\max(y_1^*, y_2^*, \ldots, y_k^*) - \min(y_1^*, y_2^*, \ldots, y_k^*)} \tag{2.6}$$

gives the scores with $\min(x_1^*, x_2^*, \ldots, x_k^*) = 0$ and $\max(x_1^*, x_2^*, \ldots, x_k^*) = 1$ and $r_{\max} = r(x_1^*, x_2^*, \ldots, x_k^*) = r(y_1^*, y_2^*, \ldots, y_k^*)$.

The scores $z_1^*, z_2^*, \ldots, z_k^*$ that minimize $r(x_1, x_2, \ldots, x_k)$ over all scores $x_1, x_2, \ldots, x_k$ consistent with the partial order are found first by finding y_i^*, that are the generalized isotonic regression of $m_i/(m_i + n_i)$ with weights $(m_i + n_i)$ and are the minimizing solution of

$$\sum_{i=1}^{k} \left(\frac{m_i}{m_i + n_i} - y_i \right)^2 (m_i + n_i) \tag{2.7}$$

over all scores y_i's consistent with the partial order.

Then the transformation of (2.6) gives the desired scores z_i^*.

Case 2: *Population* 1 *data are stochastically larger with respect to* $\preccurlyeq$ *than population* 0 *data.* For this case, the scores x_i^*, $i = 1, 2, \ldots, k$ which maximize $r(x_1, x_2, \ldots, x_k)$ over all scores x_i consistent with the partial order with $\min(x_1^*, x_2^*, \ldots, x_k^*) = 0$ and $\max(x_1^*, x_2^*, \ldots, x_k^*) = 1$, are given by (2.5) and (2.6).

The nondegenerate scores $z_1^*, z_2^*, \ldots, z_k^*$ that minimize the correlation coefficient $r(x_1, x_2, \ldots, x_k)$ over all possible nondegenerate scores x_i's consistent with the partial order correspond to one of the extreme points of the set

$$\begin{aligned} S = \{&(x_1, x_2, \ldots, x_k)\ 0 \leqslant x_i \leqslant 1,\ x_i\text{'s nondegenerate, and} \\ &\text{consistent with the desired partial order } \preccurlyeq\}. \end{aligned} \tag{2.8}$$

Let $\mathscr{U}^* = \{U_1, U_2, \ldots, U_s\}$ be the class of all nontrivial upper sets of the partial order $\preccurlyeq$. Then, the extreme points of the set S are given by p_i, $i = 1, \ldots, s$, where $p_i = (x_1, x_2, \ldots, x_k)$, and x_j, $j = 1, \ldots, k$ is defined by

$$x_j = \begin{cases} 1, & L_j \in U_i, \\ 0, & \text{otherwise}, \end{cases} \tag{2.9}$$

where $i = 1, 2, \ldots, s$ and where $U_i \in \mathscr{U}^*$.

Thus, to find the $r_{\min}$ for this case, we need to compute r for all the extreme points of S, and an extreme point where r takes the minimum value gives desired z_i^*'s.

Case 3: *Population* 0 *data are stochastically larger with respect to* $\preccurlyeq$ *than the population* 1 *data.* For this case, x_i^*'s that maximize r are given by one of the extreme points of S. The scores z_i^*'s that minimize r are given by transformation (2.6) applied to the solution of (2.7).

Below we give the minimum lower sets algorithm needed to solve (2.5). The algorithm simplifies for some special ordering for certain data examples. The solution of (2.7) can be found analogously.

To solve (2.5) by the Minimum Lower Sets Algorithm, let Av(B) denote the weighted average

$$\mathrm{Av}(B) = \frac{\sum_{\{i;L_i\in B\}} n_i}{\sum_{\{i;L_i\in B\}}(m_i+n_i)}. \tag{2.10}$$

Let $\mathscr{L} = \{B_1, B_2, \ldots, B_t\}$ denote the family of all lower sets. And let

$$\mathscr{L}_1 = \{B \in \mathscr{L} \text{ such that } \mathrm{Av}(B) \leqslant \mathrm{Av}(B_i) \text{ for all } i = 1, 2, \ldots, t\}.$$

Let $C_1 = \bigcup_{B\in\mathscr{L}_1} B$. Since the union of lower sets is again a lower set, C_1 is the largest lower set satisfying $\mathrm{Av}(C_1) \leqslant \mathrm{Av}(B_i)$ for all $i = 1, 2, \ldots, t$. Let C_1 consist of levels $L_{i_1}, L_{i_2}, \ldots, L_{i_{t_1}}$, i.e.

$$C_1 = \{L_{i_1}, L_{i_2}, \ldots, L_{i_{t_1}}\}.$$

Then

$$y_i^* = \mathrm{Av}(C_1) \quad \text{for all } i = i_1, i_2, \ldots, i_{t_1}.$$

Now let $\mathscr{L}_2$ be the class of lower sets with C_1 subtracted, i.e.

$$\mathscr{L}_2 = \{D_1, D_2, \ldots, D_t\},$$

where $D_i = B_i - C_1$, $i = 1, 2, \ldots, t$; and

$$\mathscr{L}_3 = \{D \in \mathscr{L}_2, \text{ such that } \mathrm{Av}(D) \leqslant \mathrm{Av}(D_i) \text{ for all } i = 1, 2, \ldots, t\}.$$

Set

$$C_2 = \bigcup_{D\in\mathscr{L}_3} D = \{L_{j_1}, L_{j_2}, \ldots, L_{j_{t_2}}\}.$$

Then

$$y_i^* = \mathrm{Av}(C_2) \quad \text{for all } i = j_1, j_2, \ldots, j_{t_2}.$$

The procedure is continued until we obtain scores y_i^* for all the levels $L_1, L_2, \ldots, L_k$. For a reasonably sized partial order, this is a very computationally intensive procedure.

The Minimum Lower Sets Algorithm becomes simplified for some of the special partial orderings given below:

(a) *There are p categories completely ordered, and k–p categories isolated*: Let $L_1, L_2, \ldots, L_p$ be completely ordered with $L_1 \preccurlyeq L_2 \preccurlyeq \cdots \preccurlyeq L_p$ and $L_{p+1}, \ldots, L_k$ be isolated, i.e., none of these $k-p$ categories is comparable to any of the other categories. Then $y_1^*, y_2^*, \ldots, y_p^*$ which minimize (2.5) are given by the usual isotonic regression of $n_i/(m_i+n_i)$ with weights (m_i+n_i), $i = 1, 2, \ldots, p$ (which can be conveniently found by the Pool Adjacent Violators Algorithm (PAVA) or by the Recursive Formulas of Puri and Singh, 1990) and $y_j^* = n_j/(m_j+n_j)$, $j = p+1, \ldots, k$.

Note that for $p=0$ (all categories are nominal), the statistic $(m+n)r^2_{\max}(x_1, x_2, \ldots, x_k)$, where the max is taken over all possible unrestricted scores x_i's is identical to the Pearson chi-squared test statistic of independence (see Haberman, 1981; O'Neill, 1978; Gautam and Kimeldorf, 1999).

(b) *The simple tree (star-shaped order)*: This is the order of the form $L_1 \preccurlyeq L_i$, $i = 2, 3, \ldots, k$ with no relationships between L_i and L_j for $i, j \geqslant 2$.

If $n_i/(m_i+n_i) \geqslant n_1/(m_1+n_1)$ for all $i=2,3,\ldots,k$, then $y_i^*=n_i/(m_i+n_i),\ i=1,2,\ldots,k$ minimize (2.5). If not, arrange the values $n_2/(m_2+n_2),\ldots,n_k/(m_k+n_k)$ in nondecreasing order. By relabelling the categories 2 to k, if necessary, we can assume that $n_i/(m_i+n_i)$ are nondecreasing from $i=2,3,\ldots,k$. Next find the smallest positive integer j, such that

$$A_j = \frac{\sum_{i=1}^{j} n_i}{\sum_{i=1}^{j}(m_i+n_i)} < \frac{n_{j+1}}{m_{j+1}+n_{j+1}}.$$

Such an integer exists unless $\sum_{i=1}^{j-1} n_i / \sum_{i=1}^{j-1}(m_i+n_i) \geqslant n_j/(m_j+n_j)$ for $j=2,\ldots,k$ and in this case we set $j=k$. Denote the resulting integer by j^*. Then

$$y_1^* = \frac{\sum_{i=1}^{j^*} n_i}{\sum_{i=1}^{j^*}(m_i+n_i)}$$

and the values $y_i^*=y_1^*,\ i=1,2,\ldots,j^*$ and $y_i^*=n_i/(m_i+n_i),\ i=j^*+1,\ldots,k$. (To express the values in terms of the original models, the relabelling which ordered $n_2/(m_2+n_2),\ldots,n_k/(m_k+n_k)$ needs to be taken into account.)

(c) *Matrix ordering*: If the levels of the categories are of the type (R_i,S_j), where $(R_i,S_j) \preccurlyeq (R_l,S_m)$, if $i \leqslant l, j \leqslant m$, then these are said to have matrix ordering. There is an iterative algorithm available to solve (2.5) for such an ordering if the number of categories is large (Block et al., 1994).

(d) *Separable partial order*: This is a case, arising in practice, where the partial order "separates" into disjoint partially ordered sets where the optimization issues are more manageable. Let

$$X = \{L_1,L_2,\ldots,L_k\} = \bigcup_{i=1}^{r} X_i$$

such that X_i's are pairwise disjoint and for any $L_t \in X_i$ and $L_s \in X_j,\ i \neq j, L_t$ and L_s are not comparable with respect to the partial order $\preccurlyeq$. We call such a case a *separable partial order*. Eq. (2.5) in this case reduces to minimizing

$$\sum_{j=1}^{r}\sum_{L_i\in X_j}\left[\frac{n_i}{m_i+n_i}-y_i\right]^2(m_i+n_i),$$

which is equivalent to minimizing

$$\sum_{L_i\in X_j}\left[\frac{n_i}{m_i+n_i}-y_i\right]^2(m_i+n_i)$$

for each $j=1,2,\ldots,r$. It follows that the y_i^* for $L_i \in X_j$ are given by the isotonic regression of $n_i/(m_i+n_i), L_i \in X_j$ with weights $(m_i+n_i), L_i \in X_j$ with respect to the partial orders $\preccurlyeq$ restricted to X_j.

The rescaling described by (2.6) for separable partial orders utilizes the minimum and maximum scores taken over all values of the isotonic regressions and across the separated isotonic regressions. For Cases 2 and 3, where one needs to examine extreme points described in (2.9), the separable case provides a modest amount of simplification. Basically an extreme point for the entire separable partial order must be composed from choices of extreme points of each of the separated partial orders. For instance, consider $X_1 = \{L_1, \dots, L_{k_1}\}$ and $X_2 = \{L_{k_1+1}, \dots, L_k\}$. Let U_1 be an upper set in X_1 and U_2 be an upper set in X_2. If $p_1 = (x_1, \dots, x_{k_1})$ and $p_2 = (x_{k_1+1}, \dots, x_k)$ are extreme points of S_1 and S_2, respectively, corresponding to the separated partial orders, then $p = (x_1, \dots, x_{k_1}, x_{k_1+1}, \dots, x_k)$ is an extreme point corresponds to the entire partial order.

The special partial ordering considered in (a) is a particular case of a separable partial order, where it is possible to identify the actual solution of y_i^*. A further application of this notion of separability is considered in the social mobility example in Section 3.

When a computerized algorithm for isotonic regression on partial orders is available, the following theorem (whose proof is given in Appendix A) is useful in checking whether or not the data from one population are stochastically larger with respect to the partial order $\preccurlyeq$ than the data from the other population.

Theorem 2.2. *The data from Population* 1 (0) *are stochastically larger with respect to the partial order* $\preccurlyeq$ *than the data from population* 0 (1) *if and only if the solution to* (2.7) ((2.5)) *is the constant solution*

$$y_i^* = \frac{m}{m+n} \left(\frac{n}{m+n} \right) \quad \text{for all } i = 1, 2, \dots, k.$$

3. Examples

In this section we consider several examples to illustrate the type of results one would obtain using these techniques to investigate data. The first example we consider arises from randomized clinical trials where there are no primary response variables that can be measured objectively. For trials of this nature, which are typically double-blinded, it is not uncommon to employ subjective categorical ordinal measures. Typically both the physician-investigator and the patient at the end of the trial are asked to evaluate on a multi-point ordinal scale the patient's response to treatment in comparison to their condition at the start of the trial. Such ratings are usually termed global ratings. In the analysis of such data, treatments are usually compared by evaluating separately the physician's and the patient's judgements. Due to the complexity of such an analysis, it is less usual to see an analysis which compares treatments based simultaneously upon both physician's and patient's global ratings.

In Table 2, we provide prototypical data arising from a balanced randomized double-blinded study comparing an experimental treatment to a control treatment. These data are a "variation" of the data from a clinical trial analyzed by one of the authors; the study details and actual data cannot be presented due to proprietary reasons. The physician's global ratings are ordered: "deterioration" $\preccurlyeq$ "no change" $\preccurlyeq$ "improved" $\preccurlyeq$ "substantially improved" $\preccurlyeq$ and the patient's global rating is similarly ordered. These data are considered to have arisen from 400 patients being randomized equally to the two treatments and having all patients complete the study.

The purpose of this trial was to demonstrate the efficacy (and safety) of the experimental treatment with respect to the control. In this instance, the control was thought

Table 2
Physician's global rating and patient's global rating for experimental and control treatments (Numbers of Patients)

Physician's global rating	Patient's global Worse	No difference	Better	Much better	Total
Control treatment					
Deterioration	7	48	38	1	94
No change	6	17	33	6	62
Improved	1	10	21	6	38
Substantially improved	0	0	3	3	6
Total	14	75	95	16	200
Experimental treatment					
Deterioration	5	9	13	9	36
No change	4	11	35	14	64
Improved	1	11	45	15	72
Substantially improved	0	3	14	11	28
Total	10	34	107	49	200

to be an inert treatment and it was necessary to demonstrate superiority of the experimental compared to the control. In the "proprietary" analysis of these data, only the marginal relative frequencies were compared leading to separate analyses for each of the physician's global and patient's global. There are approaches to simultaneously use both ratings in comparing the experimental to the control based upon models for repeated ordinal data, among others. Our approach to using the combined data conceptually assigns 16 scores $x_{1,1},\ldots,x_{4,4}$ to all combinations of physician's and patient's globals. We require the scores to preserve the matrix partial ordering generated by the marginal ordering. Thus, the score for (No change, Better) is required to be less than or equal to the score for (Improved, Much better), while the scores assigned to (No change, Better) and (Improved, No difference) are not required to be ordered.

To check the nature of the possible stochastic ordering for these two samples, we must in theory consider $\binom{8}{4}$ upper sets (see Sampson and Whitaker, 1988, Corollary 3.1). However, Theorem 2.2 provides a quick solution. Using the IBCR algorithm of Block et al. (1994) to solve (2.7), we found the optimum solution to be the constant function (=0.5), which shows that the experimental data are stochastically larger than the control data with respect to the given partial order. Thus, these data fall in Case 2 discussed in Section 2. The scores y_i^*, $i=1,2,\ldots,16$ which maximize $r(x_1,x_2,\ldots,x_k)$ over all scores consistent with the matrix order on physician's and patient's global ratings are given by (2.5). The IBCR algorithm of Block et al. (1994) was utilized again to compute the scores y_i^* and transformation (2.6) was applied to y_i^*. The resulting maximizing scores x_i^* are tabulated in Table 3(A).

Using in (2.2) the scores x_i^* in Table 3(A), we have $r_{\max}=0.429$ and (2.1) gives $t_{\max}=9.474$. The minimizing extreme point is given in Table 3(B) with $r_{\min}=0.029$ and $t_{\min}=0.585$.

Table 3
For randomized clinical trials data in Table 2: (A) max scores; (B) min scores; (C) "rounded" max scores and (D) max scores for extreme points

Physician's global rating	Patient's global Worse	 No difference	 Better	 Much better
(A) *Max scores*				
Deterioration	0	0	0.084	0.874
No change	0.309	0.309	0.502	0.874
Improved	0.479	0.517	0.772	0.874
Substantially improved	*	1	1	1
(B) *Min scores*				
Deterioration	0	1	1	1
No change	1	1	1	1
Improved	1	1	1	1
Substantially improved	1	1	1	1
(C) "*Rounded*" *max scores*				
Deterioration	0	0	0	1
No change	1/2	1/2	1/2	1
Improved	1/2	1/2	1/2	1
Substantially improved	1	1	1	1
(D) *Max scores for extreme points*				
Deterioration	0	0	0	1
No change	0	0	1	1
Improved	1	1	1	1
Substantially improved	1	1	1	1

*Both of the data tables in Table 2 have count 0 in the fourth row and first column. Thus, the maximizing score for this level is not unique. Any number lying between 0.479 and 1 is a max score for this level.

The fact that $t_{\min}$ is less than typical critical t-values and $t_{\max}$ is larger indicates there are some consistent scorings for which there is no treatment effect and there are other consistent scorings which yield a treatment effect. KSW (1992) term this the straddling case and note that there is a need to interpret the optimizing scores from the point of view of clinical meaningfulness.

To begin to interpret the scores in Table 3(A) we suggest using a data analytic approach useful in principle components analyses, where one "rounds" scores and interprets results based upon these. We suggest the "rounded" version given in Table 3(C).

The highest weight in Table 3(C) indicates that the best response is when *either* the physician or patient considers the patient much improved. When neither the physician nor patient sees the patient as much improved, the physicians judgement seems to be overriding in this situation—a situation where improvement is not pronounced. In this case, the physician scores judge deterioration to be a 0, and see no change or improved as virtually indistinguishable and score these halfway between deteriorated and much improved. Note that for this situation of nonpronounced improvement, given the physician categorization, the patients categorization is nonmaterial.

Table 4
British mobility data (3500 father–son data values)

Father's occupational status[a]	Son's occupational status					
	$S1$	$S2$	$S3$	$S4$	$S5$	Total
$S1$	50	45	8	18	8	129
$S2$	28	174	84	154	55	495
$S3$	11	78	110	223	96	518
$S4$	14	150	185	714	447	1510
$S5$	3	42	72	320	411	848
Total	106	489	459	1429	1017	3500

[a] Note: Status $S1$ is professional, and high administrative; status $S2$ is managerial, executive and higher grade supervisory; status $S3$ is lower grade supervisory; status $S4$ is skilled manual; and status $S5$ is semi-skilled and unskilled manual.

Table 5
Danish mobility data (2391 father–son data values)

Father's occupational status	Son's occupational status					
	$S1$	$S2$	$S3$	$S4$	$S5$	Total
$S1$	18	17	16	4	2	57
$S2$	24	105	109	59	21	318
$S3$	23	84	289	217	95	708
$S4$	8	49	175	348	198	778
$S5$	6	8	69	201	246	530
Total	79	263	658	829	562	2391

For this example, the min scores are an extreme point and hence do not need rounding. The scorings that provide the least "difference" between treatment and experimental is a dichotomy which says the patient is scored 0 if that patient is seen to be worsened or deteriorated by both the physician and the patient themselves. Otherwise, the patient is scored 1. It would seem that these minimizing scores are not anywhere as clinically meaningful as the maximizing scores.

In the spirit of finding the consistent dichotomization of categories that maximizes the 2×2 χ^2 (see Halpern, 1982), we computed the maximum of $r(x_1,\dots,x_n)$ over all 69 extreme points (i.e., 0–1 scores that are consistent with the partial order). This max occurs at the values given in Table 3D with the maximized t being 8.81. There is a fair amount of similarity between the results of Table 3(D) and those corresponding to Tables 3(A) and (C).

Another example we consider is the classic British and Danish social mobility data (Bishop et al., 1975), also Sampson and Whitaker (1989, Table 2). These data present for Britain and for Denmark, a sample of father and son pairs which are cross-classified by an ordinal social strata classification (Tables 4 and 5).

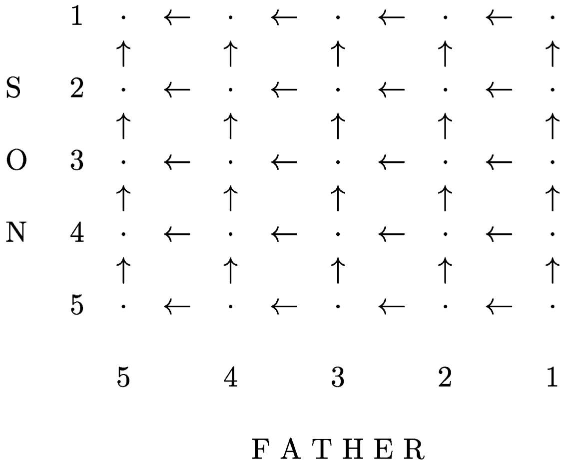

Fig. 1. Graphical representation of the upward social mobility ordering (one-step relationships, reflexive relationships not shown) arrow points to a higher order level.

For this data set, the partial orders on the categories depend on the mobility comparisons of interest. We denote the levels by (i,j) where the first coordinate denotes the father's occupational status S_i and the second, the son's status S_j, $1 \leqslant i,j \leqslant 5$.

If we are interested in upward mobility in status, we suggest the partial order (ignoring reflexive relations) which has $(a,b) \succ (c,d)$ if $(a-b) > (c-d)$. Specifically, if the change in father to son is larger, more mobility is indicated. For instance, $(S5,S3)$ indicates more mobility than $(S3,S2)$; and $(S3,S4)$ indicates more mobility than $(S1,S3)$, albeit, in this latter case, mobility in both pairs decreases, but less so in the first than the second. Thus in comparing the British and Danish populations, we seek scores for the cross-categorizations which are consistent with this partial ordering. This ordering is depicted graphically in Fig. 1. Clearly, the uppersets that are used in our notions must be consistent with this ordering. For this upward mobility ordering, it can be shown that the two data sets are not stochastically comparable. From Fig. 1 we can see $U_1 = \{(S5,S1)\}$ and $U_2 = \{(S5,S1),(S5,S2)\}$ are both upper sets. For the British data, the left-hand side of (2.4) for U_1 and U_2 become 3/3500 and 45/3500, respectively; while for the Danish data the right-hand side of (2.4) for U_1 and U_2 are 6/2391 and 14/2391, respectively. Thus, (2.4) is satisfied for U_1 and violated for U_2.

An entirely different partial ordering would be interesting if we were to consider mobility upward, as well as downward, with upward and downward mobility not comparable, and also focus only on comparing sons' status if the fathers' status were the same, i.e., we are controlling for the home environment. In this case $(a,b) \succ (c,d)$ if $a = c, b < d \leqslant c$, or, if $a = c, c \leqslant d < b$. The corresponding partial order is depicted in Fig. 2. In this case the partial order is a separable order with the separated orders being string-orders or tree-orders. In this instance, we can again demonstrate stochastic incomparability by considering the two upper sets $U_1 = \{(S1,S5)\}$ and $U_2 = \{(S2,S1)\}$.

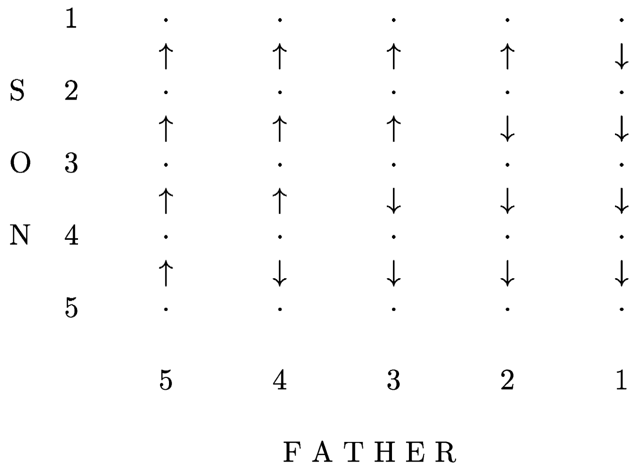

Fig. 2. Graphical representation of social mobility ordering controlling for father's status (one-step relationships, reflexive relationship not shown).

One can define a number of other partial orderings that could be used depending on the mobility notion in which one is interested. For example, one could look at (i) mobility controlling for son's status, (ii) upward mobility controlling for either son's or father's status (this reduces to separable string-orders) and (iii) a mobility version of the first partial order considered. Sampson and Whitaker (1987) examined these data from the notion of the standard matrix order, again showing stochastic incomparability.

The computation of optimizing scores, again, requires computation algorithms such as the IBCR, although the separable orders of the second instance can be handled more simply. As the data sets from the two populations are not stochastically ordered with respect to the separable partial order of the second instance depicted in Fig. 2, the optimization falls into Case 1 discussed in Section 2. The Minimum Lower Sets Algorithm was used to solve (2.5) and (2.7). Transformation (2.6) applied to the resulting solutions yielded maximizing scores x_i^* and minimizing scores z_i^* given in Table 6, where the minimizing scores, z_i^*, are given in the parentheses.

Using the optimum scores in (2.2) and (2.1), we find $r_{\max} = 0.226, t_{\max} = 17.80$; and $r_{\min} = -0.194$, $t_{\min} = -15.18$.

4. Discussion

Gautam (1991) suggests using Hotelling's T^2-statistic and treating the data multivariately. This leads to separate scores on the marginal levels. For example, consider his approach applied to the data of Table 2, which he suggests analyzing by computing the Hotelling T^2-statistic for the bivariate observations (physician's global, patient's global) by giving an increasing set of scores $u_1 \leqslant \cdots \leqslant u_4$ to the physician's rating and a set of increasing scores $v_1 \leqslant \cdots \leqslant v_4$ to the patient's ratings, and then

Table 6
Max (min) scores for social mobility data for partial order shown in Fig. 2

Father's occupational status	Son's occupational status				
	S_1	S_2	S_3	S_4	S_5
S_1	0.733	0.733	0.733	1	1
	(0)	(0.022)	(0.311)	(0.311)	(0.311)
S_2	0.504	0.504	0.504	0.834	0.836
	(0.479)	(0.270)	(0.309)	(0.309)	(0.309)
S_3	0.332	0.332	0	0.429	0.429
	(1)	(0.752)	(0.752)	(0.752)	(0.752)
S_4	0.869	0.869	0.664	0.664	0.778
	(0.326)	(0.326)	(0.326)	(0.134)	(0.134)
S_5	0.909	0.909	0.621	0.621	0.621
	(0.978)	(0.336)	(0.336)	(0.294)	(0.265)

examining optima of T^2. Combining Gautam's (1991) approach with the standard linear equivalence result (see Morrison, 1990, Section 4.2), we have that

$$\max_{\substack{u_1 \leqslant \cdots \leqslant u_4 \\ v_1 \leqslant \cdots \leqslant v_4}} T^2 = \max_{\substack{u_1 \leqslant \cdots \leqslant u_4 \\ v_1 \leqslant \cdots \leqslant v_4}} \max_{\mathbf{a} \neq 0} t^2_{\mathbf{u},\mathbf{v}}(\mathbf{a}), \tag{4.1}$$

where $t^2_{\mathbf{u},\mathbf{v}}(\mathbf{a})$ is the two-sample univariate squared t-statistic with scores $a_1 u_i + a_2 v_j$ assigned to the (i,j) cross categorization in Table 2. To compare this calculation to ours, note that the right-hand side of (4.1) can be rewritten as $\max_f t^2(f)$ where f assigns to the cross-categorization (i,j) the value $a_1 u_i + a_2 v_j$. Clearly for $\mathbf{a} \neq \mathbf{0}, f$ does not necessarily preserve the underlying partial order on Table 2. Even if one were to restrict $a_1 \geqslant 0$, $a_2 \geqslant 0$, the resulting functions then would be a strict subset of all functions preserving the partial order. In our view, Gautam's procedure should be seen as dealing with marginal monotone scores, where a_1, a_2 reflect the stochastic dependence between marginals. On the other hand, our approach is to choose scores that are monotone with respect to the joint order and reflect stochastic dependence in that sense. These linearity notions are obviously not particular to our setting. For complex partial orders corresponding to higher dimensional multivariate data, it becomes increasingly difficult to implement the scoring we propose, and optimization over a suitable subset of increasing scores is needed for feasibility. In situations akin to this, various additive models have been considered (e.g., Bacchetti, 1989).

Appendix A

Proof of Theorem 2.1. First let us assume that $r(x_1,x_2,\ldots,x_k) \geqslant 0$ for all the scores $x_1,x_2,\ldots,x_k$ consistent with the partial order $\preccurlyeq$. We need to show that population 1 data are stochastically larger with respect to $\preccurlyeq$ than population 0 data.

Since $r \geqslant 0$, this implies by (2.2) that

$$\frac{1}{n}\sum_{i=1}^{k} n_i x_i \geqslant \frac{1}{m}\sum_{i=1}^{k} m_i x_i \tag{A.1}$$

for all the scores x_i consistent with the partial order $\preccurlyeq$.

For any upper set U, let

$$x_j = \begin{cases} 1, & x_j \in U, \\ 0, & \text{otherwise}, \end{cases}$$

where $j = 1,2,\ldots,k$.

Clearly $x_1,x_2,\ldots,x_k$ are consistent with the partial ordering. Hence, (A.1) implies

$$\frac{1}{n}\sum_{\{i:L_i\in U\}} n_i \geqslant \frac{1}{m}\sum_{\{i:L_i\in U\}} m_i, \tag{A.2}$$

that is, population 1 data are stochastically greater with respect to $\preccurlyeq$ than population 0 data.

Conversely let population 1 data be stochastically greater with respect to $\preccurlyeq$ than the population 0 data, i.e., (A.2) holds.

Let $f(L_i)=x_i$ be a real valued function on $X=\{L_1,L_2,\ldots,L_k\}$ isotonic with respect to the given partial order on X. Since r is location and scale invariant, without loss of generality we assume that $x_i \geqslant 0$.

Let us define the two probability distributions on X:

$$p\{L_i\} = \frac{m_i}{m}, \quad q\{L_i\} = \frac{n_i}{n}, \quad i = 1,2,\ldots,k.$$

From (A.2), it follows that

$$P_p(U) \leqslant P_q(U) \quad \text{for all the upper sets } U. \tag{A.3}$$

Then it follows that

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^{k} m_i x_i &= \int_0^\infty P_p\{f > a\}\,\mathrm{d}a \\ &\leqslant \int_0^\infty P_q\{f > a\}\,\mathrm{d}a = \frac{1}{n}\sum_{i=1}^{k} n_i x_i, \end{aligned}$$

where the inequality follows from (A.3) and the fact that $\{f > a\}$ in this situation is an upper set. Hence, from (2.2) we have $r(x_1,x_2,\ldots,x_k) \geqslant 0$.

Proof of Theorem 2.2. Let the solution of (2.7) be given by

$$y_i^* = \frac{m}{m+n}.$$

It follows from Robertson et al. (1988, (1.3.8)) that $y_i^* = m/(m+n)$ satisfies

$$\sum_{i=1}^{k} \left(\frac{m}{m_i + n_i} - \frac{m}{m+n} \right) y_i(m_i + n_i) \leqslant 0 \tag{A.4}$$

for all scores y_i consistent with the partial order $\preccurlyeq$. Eq. (A.4) thus yields

$$\frac{1}{m} \sum_{i=1}^{k} m_i y_i \leqslant \frac{1}{n} \sum_{i=1}^{k} n_i y_i \tag{A.5}$$

for all y_i's isotonic with respect to $\preccurlyeq$, from which it follows:

$$\frac{1}{m} \sum_{\{i:L_i \in U\}} m_i \leqslant \frac{1}{n} \sum_{\{i:L_i \in U\}} n_i$$

for any upper set U, that is, the data of population 1 are stochastically larger with respect to partial order $\preccurlyeq$ than the data of population 0.

Conversely, if the data of population 1 are stochastically larger with respect to the partial order $\preccurlyeq$, than that of population 0, then it is easily seen that $y_i^* = m/(m+n)$ satisfies (1.3.7) and (1.3.8) of Robertson et al. (1988), implying that $y_i^* = m/(m+n)$ solves (2.7).

The following theorems give the optimum scores under the different scenarios. The proofs of these theorems follow similar lines as the corresponding theorems of KSW (1992) and, thus, are omitted.

Theorem A.1. *If the population* 1 *data are not stochastically smaller with respect to* $\preccurlyeq$ *than the population* 0 *data, then the scores* $y_1^*, y_2^*, \ldots, y_k^*$ *which maximize* $r(x_1, x_2, \ldots, x_k)$ *over all possible scores* x_i*'s consistent with the partial order are given by the generalized isotonic regression of* $n_i/(m_i + n_i)$ *with weights* $(m_i + n_i)$ *and the generalized isotonic regression is with respect to the partial order* $\preccurlyeq$.

Theorem A.2. *Suppose that the population* 1 *data are not stochastically larger with respect to* $\preccurlyeq$ *than the population* 0 *data. Then* $y_1^*, y_2^*, \ldots, y_k^*$ *which minimize* $r(x_1, x_2, \ldots, x_k)$ *over all scores* x_i*'s consistent with the partial order on* X *are given by the generalized isotonic regression of* $m_i/(m_i + n_i)$ *with weights* $(m_i + n_i)$.

Theorem A.3. *Suppose that the population* 1 *data are stochastically smaller with respect to* $\preccurlyeq$ *than the population* 0 *data. Then scores* $(x_1^*, x_2^*, \ldots, x_k^*)$ *which maximize* $r(x_1, x_2, \ldots, x_k)$ *over all possible nondegenerate scores consistent with the partial order are given by one of the extreme points of the set* S *defined in* (2.8). *The extreme points of* S *are given by* $p_i = (x_1, \ldots, x_k)$, $i = 1, 2, \ldots, s$, *where* x_j*'s are given by* (2.9).

Theorem A.4. *Suppose that the population* 1 *data are stochastically larger with respect to* $\preccurlyeq$ *than the population* 0 *data. Then* $(z_1^*, z_2^*, \ldots, z_k^*)$ *that minimize* $r(x_1, x_2, \ldots, x_k)$ *over all possible nondegenerate scores consistent with the partial order is one of the extreme points of the set* S *defined above.*

References

Agresti, A., 1984. Analysis of Ordinal Categorical Data. Wiley, New York.

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Bacchetti, P., 1989. Additive isotonic models. J. Amer. Statist. Assoc. 84, 289–294.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference Under Order Restrictions. Wiley, London.

Bishop, Y.M., Fienberg, S.E., Holland, P.W., 1975. Discrete Multivariate Analysis. MIT Press, Cambridge, MA.

Block, H., Qian, S., Sampson, A.R., 1994. Structure algorithms for partially ordered isotonic regressions. J. Comput. Graph. Statist. 3, 285–300.

Gautam, S.P., 1991. Applications of t, T^2 and F statistics to ordinal categorical data. Dissertation, Department of Mathematical Sciences, University of Texas at Dallas.

Gautam, S.P., Kimeldorf, G., 1999. Maximal correlation in $2 \times k$ contingency tables. Amer. Statist. 53, 336–341.

Haberman, S.J., 1981. Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. Ann. Statist. 9, 1178–1186.

Halpern, J., 1982. Maximally selected chi square statistics for small samples. Biometrics 38, 1017–1023.

Kimeldorf, G., Sampson, A.R., Whitaker, L.R., 1992. Min and max scorings for two-sample ordinal data. J. Amer. Statist. Assoc. 87, 241–247.

Marshall, A., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

Morrison, D.F., 1990. Multivariate Statistical Methods. McGraw Publishing Company, New York.

O'Neill, M.E., 1978. Distribution expansions for canonical correlations from contingency tables. J. Roy. Statist. Soc. B 40, 303–312.

Puri, P.S., Singh, H., 1990. Recursive formulas for isotonic regression useful for statistical inference under order restrictions. J. Statist. Plann. Inference 24, 1–11.

Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.

Sampson, A.R., Whitaker, L., 1988. Positive dependence, upper sets, and multidimensional partitions. Math. Oper. Res. 13, 254–264.

Sampson, A.R., Whitaker, L., 1989. Estimation of multivariate distributions under stochastic ordering. J. Amer. Statist. Assoc. 84, 549–555.