Chapter 7

The Two-Way Layout

INTRODUCTION

The procedures of this chapter are designed for statistical analyses of data collected under the auspices of an experimental design involving two factors, each at two or more levels. Our primary interest is in the relative location effects (medians) of the different levels of one of these factors, hereafter called the *treatment* factor, *within* the various levels of the second factor, hereafter called the *blocking* factor. This blocking factor is associated quite commonly with the experimental design where subjects are first divided into more homogeneous subgroups (called *blocks*) and then randomly assigned to the various treatment levels within these blocks. Such a design is called a *randomized block design*, and we will use this treatment/block terminology to describe the two-way layout setting throughout this chapter. In addition, we will refer, without loss of generality, to the k levels of a treatment as the k *treatments*. (In the case of a randomized complete block design, where the data consist of one observation on each of k treatments in each of n blocks, this represents a direct generalization of the paired replicates setting discussed in Chapter 3.)

The basic null hypothesis of interest is that of no differences in the location effects (medians) of the k treatments within each of the blocks. The alternatives considered here correspond to either general or ordered differences between the treatment effects (medians). As with the one-way layout setting in Chapter 6, we also differentiate between those cases where all of the k treatments represent "new" categories for study and those where one of the treatments corresponds to a control or a baseline category. Finally, we must deal separately with a variety of different possibilities (and correspondingly different statistical procedures) for the number of observations available from each treatment–block combination (cell), ranging from 0 (missing data), 1, to more than 1 (replications).

Sections 7.1–7.5 are devoted to the case of one observation per treatment–block cell (commonly known as a randomized complete block design). Section 7.1 presents a distribution-free test directed at general alternatives. A distribution-free test designed specifically to detect ordered differences among the k treatments is discussed in Section 7.2. Multiple comparison procedures designed to detect which, if any, treatment effects differ from one another are presented in Section 7.3 (all-treatments comparisons) and 7.4 (treatments-versus-control comparisons). In Section 7.5 we present estimators of contrasts in the treatment effects.

Nonparametric Statistical Methods, Third Edition. Myles Hollander, Douglas A. Wolfe, Eric Chicken.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

Sections 7.6–7.8 deal with settings where certain treatment–block cells yield single observations, but there are also treatment–block combinations for which we have no observations; that is, we have either zero or one observation from each treatment–block cell. Sections 7.6 and 7.7 present a distribution-free hypothesis test for general alternatives and an all-treatments multiple comparison procedure, respectively, for the structured setting where the data arise from a balanced incomplete block design (BIBD). Section 7.8 discusses a distribution-free hypothesis test for general alternatives in a two-way layout with an arbitrary configuration of either zero or one observation per cell.

In Sections 7.9 and 7.10 we discuss procedures for the setting where there is at least one observation from each cell and there are some cells with multiple observations (replications). Section 7.9 presents a distribution-free hypothesis test for general alternatives for this replications setting, with an emphasis on the special case where we have an equal number (>1) of replications in each cell. An all-treatments multiple comparison procedure for this setting of an equal number of replications is detailed in Section 7.10.

All of the procedures in Sections 7.1–7.10 are associated with within-blocks rankings (known as the *Friedman ranks*) and represent direct extensions to the two-way layout of the paired replicates sign procedures discussed in Sections 3.4–3.6. The corresponding extensions to the two-way layout of the paired replicates Wilcoxon signed ranks procedures discussed in Sections 3.1–3.3 yield asymptotically (number of blocks tending to infinity) distribution-free test and multiple comparison procedures, and we present simplified conservative versions that are nearly asymptotically distribution-free. In Sections 7.11–7.15 we discuss these extensions associated with Wilcoxon signed ranks for data from a randomized complete block design with k treatments and n blocks. Section 7.11 contains a conservative signed ranks test directed at general alternatives, and Section 7.12 presents the corresponding conservative signed ranks test procedure designed for ordered alternatives. The associated approximate signed ranks multiple comparison procedures are given in Sections 7.13 (all-treatments comparisons) and 7.14 (treatments-versus-control comparisons). Section 7.15 contains the contrast estimators linked to the Wilcoxon signed ranks.

The asymptotic relative efficiencies for translation alternatives of the procedures with respect to their normal theory counterparts are discussed in Section 7.16.

	Treatments			
Blocks	1	2	...	k
1	X_{111}	X_{121}	...	X_{1k1}
	⋮	⋮	...	⋮
	$X_{11c_{11}}$	$X_{12c_{12}}$	...	$X_{1kc_{1k}}$
2	X_{211}	X_{221}	...	X_{2k1}
	⋮	⋮	...	⋮
	$X_{21c_{21}}$	$X_{22c_{22}}$	...	$X_{2kc_{2k}}$
⋮	⋮	⋮	⋮	⋮
n	X_{n11}	X_{n21}	...	X_{nk1}
	⋮	⋮	...	⋮
	$X_{n1c_{n1}}$	$X_{n2c_{n2}}$	...	$X_{nkc_{nk}}$

Data. The data consist of $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij}$ observations, with c_{ij} observations from the combination of the ith block with the jth treatment (i.e., the (i, j)th cell), for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

Assumptions

A1. The N random variables $\{(X_{ij1}, \ldots, X_{ijc_{ij}}), i = 1, \ldots, n \text{ and } j = 1, \ldots, k\}$ are mutually independent.

A2. For each fixed (i, j), with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, k\}$, the c_{ij} random variables $(X_{ij1}, \ldots, X_{ijc_{ij}})$ are a random sample from a continuous distribution with distribution function F_{ij}.

A3. The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F(u - \beta_i - \tau_j), \quad -\infty < u < \infty, \tag{7.1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where F is a distribution function for a continuous distribution with unknown median θ, β_i is the unknown additive effect contributed by block i, and τ_j is the unknown additive treatment effect contributed by the jth treatment.

We note that Assumptions A1–A3 correspond directly to the usual two-way layout *additive* (See Comment 6) model associated with normal theory assumptions; that is, Assumptions A1–A3 are equivalent to the representation

$$X_{ijt} = \theta + \beta_i + \tau_j + e_{ijt}, \quad i = 1, \ldots, n; \quad j = 1, \ldots, k; \quad t = 1, \ldots, c_{ij},$$

where θ is the overall median, τ_j is the treatment j effect, β_i is the block i effect, and the N e's form a random sample from a continuous distribution with median 0. (Under the additional assumption of normality, the medians θ and 0 are, of course, also the respective means.)

Hypothesis

The null hypothesis of interest in Sections 7.1, 7.2, 7.6, 7.8, 7.9, 7.11, and 7.12 is that of no differences among the additive treatment effects $\tau_1, \ldots, \tau_k$, namely,

$$H_0 : [\tau_1 = \cdots = \tau_k]. \tag{7.2}$$

The null hypothesis asserts that the underlying distributions $F_{i1}, \ldots, F_{ik}$ within block i are the same, for each fixed $i = 1, \ldots, n$; that is, $F_{i1} \equiv F_{i2} \equiv \cdots \equiv F_{ik} \equiv F_i$, for $i = 1, \ldots, n$, in (7.1).

In Sections 7.1–7.5 we consider the special case of one observation per treatment–block combination (commonly known as a *randomized complete block design*), corresponding to $c_{ij} = 1$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, k$. For ease of notation in these five sections, we drop the third subscript on the X variables, since it is always equal to 1 in this setting.

7.1 A DISTRIBUTION-FREE TEST FOR GENERAL ALTERNATIVES IN A RANDOMIZED COMPLETE BLOCK DESIGN (FRIEDMAN, KENDALL-BABINGTON SMITH)

In this section we present a procedure for testing H_0 (7.2) against the general alternative that at least two of the treatment effects are not equal, namely,

$$H_1 : [\tau_1, \ldots, \tau_k \text{ not all equal}], \tag{7.3}$$

when $c_{ij} \equiv 1$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

Procedure

To compute the Friedman (1937) statistic S, we first order the k observations from least to greatest separately within each of the n blocks. Let r_{ij} denote the rank of X_{ij} in the joint ranking of the observations $X_{i1}, \ldots, X_{ik}$ in the ith block and set

$$R_j = \sum_{i=1}^{n} r_{ij} \quad \text{and} \quad R_{\cdot j} = \frac{R_j}{n}. \tag{7.4}$$

Thus, for example, R_2 is the sum (over the n blocks) of the within-blocks ranks received by the treatment 2 observations and $R_{\cdot 2}$ is the average within-blocks rank for these same observations. The Friedman statistic S is then given by

$$\begin{aligned} S &= \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(R_{\cdot j} - \frac{k+1}{2} \right)^2 \\ &= \left[\frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 \right] - 3n(k+1), \end{aligned} \tag{7.5}$$

where $(k+1)/2 = \sum_{i=1}^{n} \sum_{j=1}^{k} r_{ij}/nk$ is the average rank assigned via this within-blocks ranking scheme.

To test

$$H_0 = [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } S \geq s_\alpha; \quad \text{otherwise do not reject}, \tag{7.6}$$

where the constant s_α is chosen to make the type I error probability equal to α. The constant s_α is the upper α percentile for the null $(\tau_1 = \cdots = \tau_k)$ distribution of S. Comment 8 explains how to obtain the critical values s_α for k treatments, n blocks, and available levels of α.

Large-Sample Approximation

When H_0 is true, the statistic S has, as n tends to infinity, an asymptotic chi-square (χ^2) distribution with $k-1$ degrees of freedom. (See Comment l0 for indications of the proof.) The chi-square approximation for procedure (7.6) is

$$\text{Reject } H_0 \text{ if } S \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.7}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile point of a chi-square distribution with $k-1$ degrees of freedom. To find $\chi^2_{k-1,\alpha}$ we use the R command `qchisq`$(1-\alpha, k-1)$. For example, to find $\chi^2_{5,.05}$, we apply `qchisq(.95, 5)` and obtain $\chi^2_{5,.05} = 11.071$.

Ties

If there are ties among the k observations in a given block, assign each of the observations in a tied group the average of the within-blocks integer ranks that are associated with the tied group and compute S with these within-blocks average ranks. As a consequence of the effect that ties have on the null distribution of S, the following modification is required to apply either procedure (7.6) or the large-sample approximation in (7.7) when there are tied data values within any of the blocks. For either of these procedures, we replace S by

$$S' = \frac{12\sum_{j=1}^{k}\left(R_j - \dfrac{n(k+1)}{2}\right)^2}{nk(k+1) - [1/(k-1)]\sum_{i=1}^{n}\left\{\left(\sum_{j=1}^{g_i} t_{i,j}^3\right) - k\right\}}$$

$$= \frac{12\sum_{j=1}^{k} R_j^2 - 3n^2k(k+1)^2}{nk(k+1) - [1/(k-1)]\sum_{i=1}^{n}\left\{\left(\sum_{j=1}^{g_i} t_{i,j}^3\right) - k\right\}}, \tag{7.8}$$

where g_i denotes the number of tied groups in the ith block and $t_{i,j}$ is the size of the jth tied group in the ith block. We note that an untied observation within a block is considered to be a tied group of size 1. In particular, if there are no ties among the X's in the ith block, then $g_i = k, t_{i,j} = 1$ for each $j = 1, \ldots, k$, and the correction term for the ith block becomes $\{(\sum_{j=1}^{g_i} t_{i,j}^3) - k\} = k - k = 0$. If each block is void of ties, then we have $\sum_{i=1}^{n}\{(\sum_{j=1}^{g_i} t_{i,j}^3) - k\} = 0$ and S' (7.8) reduces to S, as given in (7.5).

We note that even the small-sample procedure (7.6) is only approximately, and not exactly, of significance level α in the presence of tied X observations within any of the blocks. To get an exact level α-test in this tied setting, see Comment 9.

EXAMPLE 7.1 ***Rounding First Base.***

The data in Table 7.1 were obtained by Woodward (1970) in a study to determine which, if any, of three methods of rounding first base is best, in the sense that it minimizes, on the average, the time to reach second base. The three methods, "round out," "narrow angle," and "wide angle" are illustrated in Figure 7.1.

Table 7.1 Rounding-First-Base Times

Players	Methods		
	Round out	Narrow angle	Wide Angle
1	5.40 (1)	5.50 (2)	5.55 (3)
2	5.85 (3)	5.70 (1)	5.75 (2)
3	5.20 (1)	5.60 (3)	5.50 (2)
4	5.55 (3)	5.50 (2)	5.40 (1)
5	5.90 (3)	5.85 (2)	5.70 (1)
6	5.45 (1)	5.55 (2)	5.60 (3)
7	5.40 (2.5)	5.40 (2.5)	5.35 (1)
8	5.45 (2)	5.50 (3)	5.35 (1)
9	5.25 (3)	5.15 (2)	5.00 (1)
10	5.85 (3)	5.80 (2)	5.70 (1)
11	5.25 (3)	5.20 (2)	5.10 (1)
12	5.65 (3)	5.55 (2)	5.45 (1)
13	5.60 (3)	5.35 (1)	5.45 (2)
14	5.05 (3)	5.00 (2)	4.95 (1)
15	5.50 (2.5)	5.50 (2.5)	5.40 (1)
16	5.45 (1)	5.55 (3)	5.50 (2)
17	5.55 (2.5)	5.55 (2.5)	5.35 (1)
18	5.45 (1)	5.50 (2)	5.55 (3)
19	5.50 (3)	5.45 (2)	5.25 (1)
20	5.65 (3)	5.60 (2)	5.40 (1)
21	5.70 (3)	5.65 (2)	5.55 (1)
22	6.30 (2.5)	6.30 (2.5)	6.25 (1)
	$R_1 = 53$	$R_2 = 47$	$R_3 = 32$

Source: W. F. Woodward (1970).

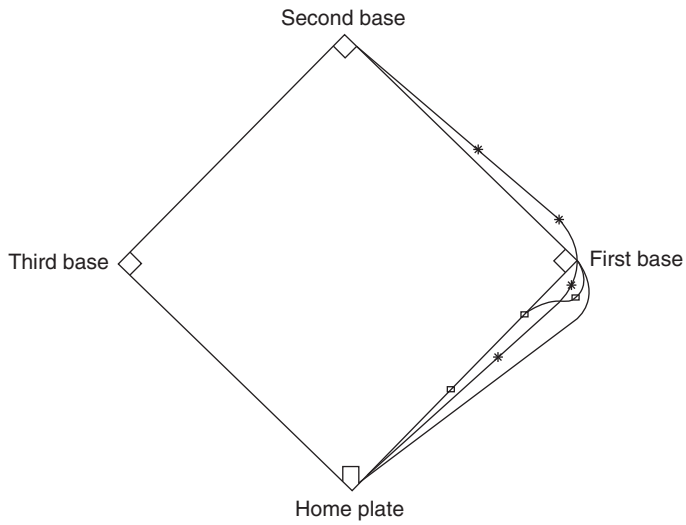

Figure 7.1 Three methods of rounding first base: ◊ path = round out method, ∗ path = narrow angle method, solid path = wide angle method.

Twenty-two baseball players participated in the study, and each of them ran from home plate to second base six times. Using a randomized order, these six trials per player were evenly divided (two each) among the three methods (round out, narrow angle, and wide angle). The entries in Table 7.1 are average times of the two runs per method from a point on the first base line 35 ft from home plate to a point 15 ft short of second base. The within-blocks (players) ranks (r_{ij}'s) of the observations are also given in Table 7.1 in parentheses after the data values (using average ranks to break the ties) and the treatment (running method) rank sums (R_1, R_2, and R_3) are provided at the bottom of the columns.

Since ties exist in blocks 7, 15, 17, and 22, we use S' (7.8). The term in braces in the denominator of (7.8) is zero for each block i in which there are no tied observations. Thus, we need to evaluate that term only for $i = 7$, 15, 17, and 22, corresponding to the blocks in which ties exist. In block 7 there is one tied group of size 2 (5.40) and one tied group of size 1 (5.35). Thus, $t_{7,1} = 2, t_{7,2} = 1$, $g_7 = 2$, and $\{(\sum_{j=1}^{g_7} t_{7,j}^3) - k\} = \{(2^3 + 1^3) - 3\} = 6$. In the same way $\{(\sum_{j=1}^{g_i} t_{i,j}^3) - k\} = 6$ for $i = 15$, 17, and 22. Hence, from (7.8) we obtain

$$S' = \frac{12[(53-44)^2 + (47-44)^2 + (32-44)^2]}{22(3)(4) - \left(\frac{1}{2}\right)(6+6+6+6)} = 11.1.$$

For the large-sample approximation, we compare the value of S' to the chi-square distribution with $k - 1 = 2$ degrees of freedom. Since `1 - pchisq (11.1, 2) = 1 -.9961 = .0039`, we see that the lowest level at which we reject H_0, using the large-sample procedure (7.7) adjusted for ties, is approximately .004. Hence, there is strong evidence here to reject the hypothesis that the methods are equivalent with respect to time to reach second base.

Comments

1. *Basic Model*. Model (7.1) is the most basic form of the two-way layout There is just one observation per cell, and we assume that there is no interaction between the block and treatment factors.
2. *More General Setting*. We could replace Assumptions A1–A3 and H_0 (7.2) with the more general null hypothesis that all possible $(k!)^n$ rank configurations for the r_{ij}'s are equally likely. Procedure (7.6) remains distribution-free for this more general hypothesis.
3. *Design Rationale*. The n blocks in this basic two-way layout design represent an effort to reduce experimental errors and prevent misleading comparisons of "apples and oranges." (We prefer to compare apples with apples.) Thus, in Example 7.1, the 22 blocks correspond to 22 different baseball players. The treatments are to be assigned at random within each block (i.e., in each block, the order in which each player is assigned to run the three different rounding-first-base methods should be decided by a random mechanism, where each of the six possible orders has equal probability of being chosen, and the assignments in the different blocks are to be independent). Note that in the Procedure, we rank only within each block. Thus, in block 1, for example, the three treatment times of player 1 are compared. This is an attempt to eliminate a nuisance effect due to player 1's intrinsic speed. It would be foolish to compare round out times of player 1 with wide angle times of player 2 if player 1 is a (slow) 200-1b catcher

and player 2 is a (speedy) 160-1b shortstop. In such a comparison, a difference in treatment effects would be confounded with the basic speed differences of the players, the latter being of little or no interest in this particular experiment.

4. *Motivation for the Test*. Under Assumptions A1–A3 and H_0 (7.2), each of the block rank vectors $\mathbf{R}_i^* = (r_{i1}, \ldots, r_{ik}), i = 1, \ldots, n$, has a uniform distribution over the set of all $k!$ permutations of the vector of integers $(1, 2, \ldots, k)$. It follows that

$$E_0(r_{ij}) = \frac{1}{k!}(k-1)! \sum_{t=1}^{k} t = \frac{k+1}{2},$$

the average rank being assigned separately in each of the blocks. Thus, we have

$$E_0(R_{\cdot j}) = E_0\left(\frac{1}{n}R_j\right) = \frac{1}{n}E_0\left(\sum_{i=1}^{n} r_{ij}\right) = \frac{1}{n}\sum_{i=1}^{n} E_0(r_{ij})$$
$$= \frac{n(k+1)}{2n} = \frac{k+1}{2}, \quad \text{for } j = 1, \ldots, k,$$

and we would expect the $R_{\cdot j}$'s to be close to $(k+1)/2$ when H_0 is true. Since the test statistic S (7.5) is a constant times a sum of squared differences between the observed treatment average ranks, $R_{\cdot j}$ and their common null expected value, $E_0(R_{\cdot j}) = (k+1)/2$, small values of S represent agreement with H_0 (7.2). When the τ's are not all equal, we would expect a portion of the associated treatment average ranks to differ from their common null expectation, $(k+1)/2$, with some tending to be smaller and some larger. The net result (after squaring the observed differences to obtain the $\left[R_j - (k+1)/2\right]^2$ terms) would be a large value of S. This naturally suggests rejecting H_0 in favor of H_1 (7.3) for large values of S and motivates procedures (7.6) and (7.7). (See also Comment 5.)

5. *Connection to Normal Theory Test*. The Friedman S statistic also arises naturally if we apply the usual two-way layout $\mathcal{F}$ statistic to the ranks instead of the actual observations. Then S may be written as $S = [12/k(k+1)]$ SST, where SST is the treatment sum of squares applied to the ranks.

6. *Assumptions.* We emphasize that Assumption A3 stipulates that the nk cell distributions $F_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, k$, can differ at most in their locations (medians) and that these location differences (if any) must be a result of additive block and/or treatment effects (i.e., there is no interaction between the treatment and block factors). In particular, Assumption A3 requires that the nk underlying distributions belong to the same general family (F) and that they do not differ in scale parameters (variability). We do note, that the test procedure (7.6) remains distribution-free under the less restrictive setting where Assumption A3 is replaced by the weaker condition

   A3′. The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F_i(u - \tau_j), \quad -\infty < u < \infty,$$

   for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where $F_1, \ldots, F_n$ are arbitrary distribution functions for continuous distributions with unknown medians $\theta_1, \ldots, \theta_n$,

respectively, and, as before, τ_j is the unknown additive treatment effect contributed by the jth treatment.

Assumption A3 then corresponds to Assumption A3′ with the additional condition that $F_1 \equiv \cdots \equiv F_n$. (See also Comment 2.)

7. *Special Case of Two Treatments*. For the case of $k = 2$ treatments, the procedures in (7.6) and (7.7) are equivalent to the exact and large-sample approximation forms, respectively, of the two-sided sign test, as discussed in Section 3.4.

8. *Derivation of the Distribution of S under H_0 (No-Ties Case)*. The null distribution of S (7.5) can be obtained by using the fact that under H_0 (7.2), all possible $(k!)^n$ rank configurations for the r_{ij}'s are equally likely. We now take $k = 4$, $n = 2$ to illustrate how the null distribution can be derived. In this case, S (7.5) reduces to $S = (.3R^* - 30)$, where $R^* = R_1^2 + R_2^2 + R_3^2 + R_4^2$. We note that S does not vary with changes of the names of the blocks or with relabeling of the k samples. Thus, for example,

(a)

	I	II	III	IV
Block 1	1	2	3	4
Block 2	3	1	2	4

(b)

	I	II	III	IV
Block 1	3	1	2	4
Block 2	1	2	3	4

yield the same value of S, because (b) is obtained from (a) by reversing the roles of blocks 1 and 2. Similarly,

(c)

	I	II	III	IV
Block 1	1	2	3	4
Block 2	3	1	2	4

(d)

	I	II	III	IV
Block 1	2	1	3	4
Block 2	1	3	2	4

yield the same value of S, since (d) is obtained from (c) by reversing the roles of samples I and II. Instead of $(4!)^2$ rank configurations, therefore, we list only $4! = 24$ configurations (the 24 different configurations in block 2 corresponding to a fixed configuration 1, 2, 3, 4 in block 1) and their associated values of R^* and S.

(a)

I	II	III	IV
1	2	3	4
1	2	3	4

$R^* = 120, S = 6$

(b)

I	II	III	IV
1	2	3	4
1	2	4	3

$R^* = 118, S = 5.4$

(c)

I	II	III	IV
1	2	3	4
1	3	4	2

$R^* = 114, S = 4.2$

(d)

I	II	III	IV
1	2	3	4
1	3	2	4

$R^* = 118, S = 5.4$

(e)

I	II	III	IV
1	2	3	4
1	4	2	3

$R^* = 114, S = 4.2$

(f)

I	II	III	IV
1	2	3	4
1	4	3	2

$R^* = 112, S = 3.6$

(g)

I	II	III	IV
1	2	3	4
2	1	3	4

$R^* = 118, S = 5.4$

(h)

I	II	III	IV
1	2	3	4
2	1	4	3

$R^* = 116, S = 4.8$

(i)

I	II	III	IV
1	2	3	4
2	3	4	1

$R^* = 108, S = 2.4$

(j)

I	II	III	IV
1	2	3	4
2	3	1	4

$R^* = 114, S = 4.2$

(k)

I	II	III	IV
1	2	3	4
2	4	1	3

$R^* = 110, S = 3$

(l)

I	II	III	IV
1	2	3	4
2	4	3	1

$R^* = 106, S = 1.8$

(m)

I	II	III	IV
1	2	3	4
3	1	2	4

$R^* = 114, S = 4.2$

(n)

I	II	III	IV
1	2	3	4
3	1	4	2

$R^* = 110, S = 3$

(o)

I	II	III	IV
1	2	3	4
3	2	4	1

$R^* = 106, S = 1.8$

(p)

I	II	III	IV
1	2	3	4
3	2	1	4

$R^* = 112, S = 3.6$

(q)

I	II	III	IV
1	2	3	4
3	4	1	2

$R^* = 104, S = 1.2$

(r)

I	II	III	IV
1	2	3	4
3	4	2	1

$R^* = 102, S = .6$

(s)

I	II	III	IV
1	2	3	4
4	1	2	3

$R^* = 108, S = 2.4$

(t)

I	II	III	IV
1	2	3	4
4	1	3	2

$R^* = 106, S = 1.8$

(u)

I	II	III	IV
1	2	3	4
4	2	1	3

$R^* = 106, S = 1.8$

(v)

I	II	III	IV
1	2	3	4
4	2	3	1

$R^* = 102, S = .6$

(w)

I	II	III	IV
1	2	3	4
4	3	1	2

$R^* = 102, S = .6$

(x)

I	II	III	IV
1	2	3	4
4	3	2	1

$R^* = 100, S = 0$

Thus, we find

$$P_0\{S = 6\} = \tfrac{1}{24}, \quad P_0\{S = 5.4\} = \tfrac{3}{24}, \quad P_0\{S = 4.8\} = \tfrac{1}{24},$$

$$P_0\{S = 4.2\} = \tfrac{4}{24}, \quad P_0\{S = 3.6\} = \tfrac{2}{24}, \quad P_0\{S = 3\} = \tfrac{2}{24},$$

$$P_0\{S = 2.4\} = \tfrac{2}{24}, \quad P_0\{S = 1.84\} = \tfrac{4}{24}, \quad P_0\{S = 1.2\} = \tfrac{1}{24},$$

$$P_0\{S = .6\} = \tfrac{3}{24}, \quad P_0\{S = 0\} = \tfrac{1}{24}.$$

The probability, under H_0, that S is greater than or equal to 5.4, for example, is therefore

$$\begin{aligned} P_0\{S \geq 5.4\} &= P_0\{S = 5.4\} + P_0\{S = 6\} \\ &= \tfrac{3}{24} + \tfrac{1}{24} = \tfrac{1}{6}. \end{aligned}$$

Note that we have derived the null distribution of S without specifying the common form (F) of the underlying distribution function for the X's under H_0

beyond the point of requiring that it be continuous. This is why the test procedure (7.6) based on S is called a *distribution-free Procedure*. From the null distribution of S, we can determine the critical value s_α and control the probability α of falsely rejecting H_0 when H_0 is true, and this error probability does not depend on the specific form of the common underlying continuous X distribution.

For a given number of treatments k and blocks n, the R command `cFrd(α,k,n)` can be used to find the available upper-tail critical values s_α for possible values of S. For a given available significance level α, the critical value s_α then corresponds to $P_0(S \geq s_\alpha) = \alpha$ and is given by `cFrd(α, k, n)` $= s_\alpha$. Thus, for example, for $k = 5$ and $n = 7$, we have $P_0(S \geq 10.40) = .0261$, so that $s_{.0261} =$ `cFrd(.0261, 5, 7)` $= 10.40$ for $k = 5$ and $n = 7$.

9. *Exact Conditional Distribution of S with Ties among the X Values*. To have a test with exact significance level even in the presence of tied X's, we need to consider all $(k!)^n$ block rank configurations, where now these within-blocks ranks are obtained by using average ranks to break the ties. As in Comment 8, it still follows that under H_0 each of the $(k!)^n$ block rank configurations (now with these tied ranks) is equally likely. For each such configuration, the value of S is computed and the results are tabulated. We illustrate this construction only for the very limited case of $k = 3$, $n = 2$, and the tied data $X_{11} = 2.4$, $X_{12} = 3.0$, $X_{13} = 3.0$, $X_{21} = 4.0$, $X_{22} = 6.0$, and $X_{23} = 3.0$. Using average ranks to break within-blocks ties, the observed rank vector is $(r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}) = (1, 2.5, 2.5, 2, 3, 1)$. Thus, $R_1 = 3, R_2 = 5.5, R_3 = 3.5$, and the attained value of S is

$$S = \left[\frac{12}{2(3)(4)}\{(3)^2 + (5.5)^2 + (3.5)^2\} - 3(2)(4)\right] = 1.75.$$

To assess the significance of S, we obtain its conditional null distribution by considering the 36 equally likely (under H_0) possible rank configurations (i.e., permutation combinations) of the observed rank vector (1, 2.5, 2.5, 2, 3, 1). These 36 configurations and associated values of S are as follows:

I	II	III		I	II	III	
1	2.5	2.5		1	2.5	2.5	
2	3	1	$S = 1.75$	2	3	1	$S = 1.75$
2.5	1	2.5		2.5	1	2.5	
2	3	1	$S = 0.25$	2	3	1	$S = 0.25$
2.5	2.5	1		2.5	2.5	1	
2	3	1	$S = 3.25$	2	3	1	$S = 3.25$
1	2.5	2.5		1	2.5	2.5	
2	1	3	$S = 1.75$	2	1	3	$S = 1.75$
2.5	1	2.5		2.5	1	2.5	
2	1	3	$S = 3.25$	2	1	3	$S = 3.25$
2.5	2.5	1		2.5	2.5	1	
2	1	3	$S = 0.25$	2	1	3	$S = 0.25$
1	2.5	2.5		1	2.5	2.5	
1	2	3	$S = 3.25$	1	2	3	$S = 3.25$
2.5	1	2.5		2.5	1	2.5	
1	2	3	$S = 1.75$	1	2	3	$S = 1.75$

I	II	III		I	II	III	
2.5	2.5	1		2.5	2.5	1	
1	2	3	$S = 0.25$	1	2	3	$S = 0.25$
1	2.5	2.5		1	2.5	2.5	
3	2	1	$S = 0.25$	3	2	1	$S = 0.25$
2.5	1	2.5		2.5	1	2.5	
3	2	1	$S = 1.75$	3	2	1	$S = 1.75$
2.5	2.5	1		2.5	2.5	1	
3	2	1	$S = 3.25$	3	2	1	$S = 3.25$
1	2.5	2.5		1	2.5	2.5	
1	3	2	$S = 3.25$	1	3	2	$S = 3.25$
2.5	1	2.5		2.5	1	2.5	
1	3	2	$S = 0.25$	1	3	2	$S = 0.25$
2.5	2.5	1		2.5	2.5	1	
1	3	2	$S = 1.75$	1	3	2	$S = 1.75$
1	2.5	2.5		1	2.5	2.5	
3	1	2	$S = 0.25$	3	1	2	$S = 0.25$
2.5	1	2.5		2.5	1	2.5	
3	1	2	$S = 3.25$	3	1	2	$S = 3.25$
2.5	2.5	1		2.5	2.5	1	
3	1	2	$S = 1.75$	3	1	2	$S = 1.75$

Since each of these values of S has null probability $\frac{1}{36}$, it follows that

$$P_0\{S = 0.25\} = P_0\{S = 1.75\} = P_0\{S = 3.25\} = \tfrac{1}{3}.$$

This distribution is called the *conditional distribution* or the *permutation distribution* of S, given the tied ranks $\{(1, 2.5, 2.5), (1, 2, 3)\}$. For the particular observed value $S = 1.75$, we have $P_0\{S \geq 1.75\} = \frac{2}{3}$.

10. *Large-Sample Approximation*. Define the random variables $T_j = R_{.j} - E_0(R_{.j}) = R_{.j} - (k+1)/2$, for $j = 1, \ldots, k$. Since each $R_{.j} = \sum_{i=1}^{n} r_{ij}/n$ is an average, it is not surprising (see, e.g., pages 388–389 of Lehmann (1975) for justification) that a properly standardized version of the vector $\mathbf{T}^* = (T_1, \ldots, T_{k-1})$ has an asymptotic (n tending to infinity) $(k-1)$-variate normal distribution with mean vector $\mathbf{0} = (0, \ldots, 0)$ and appropriate covariance matrix $\boldsymbol{\Sigma}$ when the null hypothesis H_0 is true. (Note that $\boldsymbol{T}^*$ does not include $T_k = R_{.k} - (k+1)/2$, because T_k can be expressed as a linear combination of $T_1, \ldots, T_{k-1}$. This is the reason that the asymptotic normal distribution is $(k-1)$-variate and not k-variate.) Since the test statistic S (7.5) is a quadratic form in the variables $(T_1, \ldots, T_{k-1})$, it is, therefore, quite natural that S has an asymptotic (n tending to infinity) chi-square distribution with $k-1$ degrees of freedom.
11. *Competitor Based on Wilcoxon Signed Ranks*. The statistic S (7.5) utilizes the treatment observations only through comparisons within blocks. As noted in Comment 7, this provides a natural extension of the sign test procedure for paired data and it is this restriction to within-blocks comparisons that leads directly to the distribution-free nature of procedure (7.6). An alternative approach would be to extend the (generally) more powerful signed rank test procedure, as discussed in Section 3.1. This approach utilizes between-block comparisons

of the observations and is discussed further in Section 7.11. The associated test procedure utilizing between-blocks signed rank comparisons is (generally) more powerful than the Friedman test based on S (7.5). However, this two-way layout signed rank procedure is no longer exactly distribution-free for small numbers (n) of blocks and tests based on this approach require the use of a large-sample approximation.

12. *Consistency of the S Test*. Replace Assumptions A1–A3 by the less restrictive Assumption $A1' : X_{ij} = \beta_i + e_{ij}$, where the e's are mutually independent, and Assumption $A2' : e_{1j}, \ldots, e_{nj}$ come from the same continuous population $\prod_j, j = 1, \ldots, k$, but where $\prod_1, \ldots, \prod_k$ are not assumed to be identical. Then the test defined by (7.6) is consistent against alternatives for which $\sum_{v=1}^{k}(1 - p_{uv}) \neq \sum_{v=1}^{k} p_{uv}$ for at least one $u \in \{1, \ldots, k\}$, where $p_{uv} = P(e_{iu} < e_{iv})$ with e_{iu} a random member from $\prod_u$ and e_{iv} a random member from $\prod_v$ that is independent of e_{iu}.

Properties

1. *Consistency*. See Noether (1967a, p. 54) and Comment 12.
2. *Asymptotic Chi-Squaredness*. See Lehmann (1975, pp. 388–389).
3. *Efficiency*. See van Elteren and Noether (1959) and Section 7.16.

Problems

1. Goldsmith and Nadel (1969) have studied respiratory function following exposure to various levels of ozone for periods of 1 h. The subjects were four presumably healthy males employed by the California State Department of Public Health. The objective measurement used was airway resistance as evaluated by the body plethysmographic technique (see DuBois et al. (1956) and Comroe, Botelho, and DuBois (1959)). Goldsmith and Nadel reported average values for four consecutive measurements taken immediately prior to and again about 5 min after termination of each level of ozone exposure. Table 7.2 is based on a subset of the Goldsmith-Nadel data, where the tabled values are average airway resistance after ozone exposure minus average airway resistance prior to ozone exposure. Use procedure (7.6) to test H_0.

2. Show that the two expressions for S in (7.5) are, indeed, equivalent.

3. Could Friedman's test be applied to data from a one-way layout in which there are the same number, n, of observations from each of the k treatments? Explain. Should Friedman's test be applied to such data? Explain.

Table 7.2 Effect of Experimental Ozone Exposures on Airway Resistance (cm H_2Ol/s)

Subject	After .1 ppm	After .6 ppm	After 1.0 ppm
1	−.08	.01	.06
2	.21	.17	.19
3	.50	−.11	.34
4	.14	.07	.14

Source: J. R. Goldsmith and J. A. Nadel (1969).

4. Show directly, or illustrate by means of an example, that the maximum value of S is $S_{\max} = n(k-1)$. For what configuration is this maximum achieved?

5. Creatine phosphokinase (CPK) is a skeletal muscle isoenzyme that is often found to be elevated in the serum of acutely psychotic subjects during the initial stages of a psychotic episode. A number of variables known to affect serum CPK activity have been evaluated as possible causes of the serum CPK activity elevations observed during acute psychotic episodes. One such variable of interest is that of physical exercise, which is well known to increase serum CPK levels in normal subjects. In this regard, Goode and Meltzer (1976) studied the relationship between isometric exercises (designed to strengthen and tone muscle without lengthening and contracting the muscles themselves) and increased CPK levels in psychotic patients. In particular, they were interested in whether the elevation of CPK in the serum of psychiatric patients may be in part due to increased covert isometric motor activity. The subjects in their study were patients hospitalized on a research unit at the Illinois Psychiatric Institute. Fourteen such patients were isometrically exercised following remission of psychotic symptoms, usually 2–4 weeks after admission. The 60-min isometric exercise procedure involved stationary wall bars and required maximal use of all major muscle groups. The subjects described the exercises as extremely fatiguing and at or near the limits of their endurances.

Table 7.3 contains the plasma CPK activity (mU/l) levels for each of these 14 patients prior to the period of isometric exercises, as well as at 18 and 42 h after completion of such exercises. Also recorded for each patient is the peak plasma CPK activity exhibited during the period of psychosis immediately following admission to the Institute.

Use these data to assess whether there are any differences in CPK activity between the four patient conditions considered in Table 7.3.

6. Suppose $k = 3$ and $n = 13$. Compare the critical region for the exact level $\alpha = .025$ test of H_0 (7.2) based on S with the critical region for the corresponding nominal level $\alpha = .025$ test based on the large-sample approximation. What is the exact significance level of this .025 nominal level test based on the large-sample approximation?

7. Suppose $k = 3$ and $n = 3$. Obtain the form of the exact null (H_0) distribution of S for the case of no-tied observations.

8. Suppose $k = 4$ and $n = 8$. Compare the critical region for the exact level $\alpha = .005$ test of H_0 (7.2) based on S with the critical region for the corresponding nominal level $\alpha = .005$

Table 7.3 Effect of lsometric Exercise on Serum Creatine Phosphokinase (CPK) Activity (mU/l) in Psychotic Patients

Subject	Preexercise	19 h postexercise	42 h postexercise	Peak-psychotic period
1	27	101	82	63
2	30	112	50	78
3	24	26	68	69
4	54	89	135	1,137
5	21	30	49	57
6	36	41	48	800
7	36	29	46	105
8	16	20	8	111
9	21	26	25	61
10	26	25	31	74
11	65	60	69	190
12	25	27	28	107
13	19	18	21	306
14	48	41	28	109

Source: D. J. Goode and H. Y. Meltzer (1976).

test based on the large-sample approximation. What is the exact significance level of this .005 nominal level test based on the large-sample approximation?

9. Suppose $k = 3$ and $n = 2$, and we observe the data $X_{11} = 3.6, X_{12} = 3.6, X_{13} = 5.2, X_{21} = 4.3, X_{22} = 5.2$, and $X_{23} = 4.3$. What is the conditional probability distribution of S under H_0 (7.2) when average ranks are used to break ties among the X's? How extreme is the observed value of S in this conditional null distribution? Compare this fact with that obtained by taking the observed value of S to the (incorrect) unconditional null distribution of S.

10. Consider the CPK activity data in Table 7.3. Ignoring the patients' peak psychotic period data, assess the conjecture that isometric exercise has an effect on the CPK activity of psychotic patients.

11. Use the CPK data in Table 7.3 and an appropriate nonparametric test procedure to assess whether there is any difference between peak CPK activity during the psychotic period and peak CPK activity over the combined pre/post exercise periods.

12. Nicholls and Ling (1982) conducted a study to assess the effectiveness of a system employing hand cues in the teaching of language to severely hearing-impaired children. In particular, they considered syllables presented to hearing-impaired children under the following seven conditions: (A) audition, (L) lip reading, (AL) audition and lip reading, (C) cued speech, (AC) audition and cued speech, (LC) lip reading and cued speech, and (ALC) audition, lip reading, and cued speech. The 18 subjects in the study were all severely hearing-impaired children who had been taught through the use of cued speech for at least 4 years. Syllables were presented to the subjects under each of the seven conditions (presented in random orders) and the subjects were asked in each case to identify the consonants in each syllable by writing down what they perceived them to be. The subjects' results were scored by marking properly identified consonants in the appropriate order as correct. After tallying the responses, an overall percentage correct was assigned to each participant under each experimental condition. These correct percentage data for the 18 children in the study are given in Table 7.4.

Table 7.4 Percentage Consonants Correctly Identified under Each of the Conditions: (A) Audition, (L) Lip Reading, (AL) Audition and Lip Reading, (C) Cued Speech, (AC) Audition and Cued Speech, (LC) Lip Reading and Cued Speech, and (ALC) Audition, Lip Reading, and Cued Speech

Subject	A	L	AL	C	AC	LC	ALC
1	1.1	36.9	52.4	42.9	31.0	83.3	63.0
2	1.1	33.3	34.5	34.5	41.7	77.3	81.0
3	13.0	28.6	40.5	33.3	44.0	81.0	76.1
4	0	23.8	22.6	33.3	33.3	69.0	65.5
5	11.9	40.5	57.1	35.7	46.4	98.8	96.4
6	0	27.4	46.4	42.9	47.4	78.6	77.4
7	5.0	20.2	22.6	35.7	37.0	69.0	73.8
8	4.0	29.8	42.9	13.0	33.3	95.2	91.7
9	0	27.4	38.0	42.9	45.2	89.3	85.7
10	1.1	26.2	31.0	31.0	32.1	70.2	71.4
11	2.4	29.8	38.0	34.5	46.4	86.9	92.9
12	0	21.4	21.4	41.7	33.3	67.9	59.5
13	0	32.1	33.3	44.0	34.5	86.9	82.1
14	0	28.6	23.8	32.1	39.3	85.7	72.6
15	1.1	28.6	29.8	41.7	35.7	81.0	78.6
16	1.1	36.9	33.3	25.0	31.0	95.2	95.2
17	0	27.4	26.1	40.5	44.0	91.7	89.3
18	0	41.7	35.7	42.9	45.2	95.2	95.2

Source: G. H. Nicholls and D. Ling (1982).

Use these data to assess whether there are any differences in the effectiveness of these seven conditions for teaching severely hearing-impaired children.

13. Consider the study with severely hearing-impaired children in Problem 12. Using the percentage correctly identified data from Table 7.4, assess whether there are any differences in the effectiveness of the three stand-alone conditions A, L, and C for teaching severely hearing-impaired children.

14. Consider the study with severely hearing-impaired children in Problem 12. Using the percentage correctly identified data for only the first eight children (and the proper correction for ties), assess whether there are any differences in the teaching effectiveness from adding one or more of the factors L (lip reading) and C (cued speech) to the baseline A (audition) approach.

7.2 A DISTRIBUTION-FREE TEST FOR ORDERED ALTERNATIVES IN A RANDOMIZED COMPLETE BLOCK DESIGN (PAGE)

In many practical two-way layout settings where an additive model is appropriate, it is also the case that the treatments are such that the appropriate alternatives to no differences in treatment effects (H_0) are those of increasing (or decreasing) treatment effects according to some natural labeling for the treatments. Examples of such settings include treatments corresponding to quality or quantity of materials, severity of disease, drug dosage levels, and intensity of stimulus. We note that the Friedman procedure (7.6) does not utilize any such partial prior information regarding the postulated alternative ordering. The statistic S (7.5) takes on the same value for all possible $k!$ labelings of the treatments. In this section, we consider a procedure for testing H_0 (7.2) against the a priori ordered alternatives,

$$H_2 : [\tau_1 \leq \tau_2 \leq \cdots \leq \tau_k, \text{ with at least one strict inequality}]. \tag{7.9}$$

The Page test of this section is preferred to the Friedman test in Section 7.1 when the treatments can be labeled a priori in such a way that the experimenter expects any deviation from H_0 (7.2) to be in the particular direction associated with H_2 (7.9). We emphasize, however, that the labeling of the treatments, so that the ordered alternatives (7.9) are appropriate, *cannot* depend on the observed sample values. This labeling must correspond completely to a factor (s) implicit in the nature of the *experimental design* and *not* the *observed data.*

Procedure

First, we must label the treatments so that they are in the expected order associated with the alternative H_2 (7.9). (This labeling must be done prior to data collection.) To compute the Page (1963) statistic L, we once again rank within blocks and compute the Friedman treatment sums of ranks $R_1, \ldots, R_k$ as defined in (7.4). The Page statistic L is then the weighted combination of these rank sums given by

$$L = \sum_{j=1}^{k} jR_j = R_1 + 2R_2 + \cdots + kR_k. \tag{7.10}$$

To test

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the ordered alternative

$$H_2 : [\tau_1 \leq \tau_2 \leq \cdots \leq \tau_k, \text{ with at least one strict inequality}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } L \geq l_\alpha; \qquad \text{otherwise do not reject,} \tag{7.11}$$

where the constant l_α is chosen to make the type I error probability equal to α. The constant l_α is the upper α percentile for the null $(\tau_1 = \cdots = \tau_k)$ distribution of L. Comment 17 explains how to obtain the critical value l_α for k treatments, n blocks, and available levels of α.

Large-Sample Approximation

The large-sample approximation is based on the asymptotic (n tending to infinity) normality of L, suitably standardized. We first need to know the expected value and variance of L when the null hypothesis is true. Under H_0, the expected value and variance of L are

$$E_0(L) = \frac{nk(k+1)^2}{4} \tag{7.12}$$

and

$$\text{var}_0(L) = \frac{nk^2(k+1)(k^2-1)}{144}, \tag{7.13}$$

respectively. These expressions for $E_0(L)$ and $\text{var}_0(L)$ are verified by direct calculations in Comment 18 for the special case of $k = 3$ and $n = 2$. General derivations of both expressions are outlined in Comment 20.

The standardized version of L is

$$L^* = \frac{L - E_0(L)}{\sqrt{\text{var}_0(L)}} = \frac{L - \left[\dfrac{nk(k+1)^2}{4}\right]}{\left\{\dfrac{nk^2(k+1)(k^2-1)}{144}\right\}^{1/2}}. \tag{7.14}$$

When H_0 is true, L^* has, as n tends to infinity, an asymptotic $N(0,1)$ distribution (see Comment 20 for indications of the proof). The normal theory approximation for procedure (7.11) is

$$\text{Reject } H_0 \text{ if } L^* \geq_{Z_\alpha}; \quad \text{otherwise do not reject.} \tag{7.15}$$

Ties

If there are ties among the k X's within any of the n blocks, assign each of the observations in a tied group the average of the integer ranks that are associated with the tied group and compute L with these average ranks.

We note that even procedure (7.11) using these average ranks to break ties and the critical value l_α is only approximately, and not exactly, of significance level α in the presence of tied X observations within any of the blocks. To get an exact level α test in this tied setting, see Comment 19. (See also Comment 21 regarding the use of the large-sample approximation in the case of within-blocks ties.)

EXAMPLE 7.2 ***Breaking Strength of Cotton Fibers.***

An experiment in Cochran and Cox (1957, p. 108) considered the effect, in terms of breaking strength of cotton fibers, of the level of potash (K_2O) in the soil. Five levels of potash were applied ($k = 5$) in a randomized block pattern with three blocks ($n = 3$). The criterion used for the analysis was the Pressley strength index, obtained by measuring the breaking strength of a bundle of fibers of a given cross-sectional area. A single sample of cotton was taken from each plot, and four determinations were made on each sample. The main entries of Table 7.5 are the means of the four determinations and the parenthetical values are the within-block ranks. (No dimensions are associated with the data of Table 7.5, because the machine that measures the strength index is calibrated in arbitrary units.)

We are interested here in using procedure (7.11) to test the hypothesis of equivalent strengths versus the ordered alternative that specifies a trend of decreasing breaking strength with increasing levels of potash. For the purpose of illustration, we take the significance level to be $\alpha = .0097$. Applying the R command `cPage(α,k,n)` with $k = 5$ and $n = 3$, we find `cPage(.0097, 5, 3) = 155`. That is, $P_0(L \geq 155) = .0097$, and we have that $l_{.0097} = 155$ and procedure (7.11) becomes

$$\text{Reject } H_0 \text{ if } L \geq 155.$$

Now, we illustrate the computations leading to the sample value of L (7.10). Using the treatment sums of within-block ranks given in Table 7.3, we see from (7.10) that

$$\begin{aligned} L &= R_1 + 2R_2 + 3R_3 + 4R_4 + 5R_5 \\ &= 5 + 2(5) + 3(9) + 4(14) + 5(12) = 158. \end{aligned}$$

Table 7.5 Strength Index of Cotton

	Potash (lb/acre)				
Replications	144	108	72	54	36
1	7.46 (2)	7.17 (1)	7.76 (4)	8.14 (5)	7.63 (3)
2	7.68 (2)	7.57 (1)	7.73 (3)	8.15 (5)	8.00 (4)
3	7.21 (1)	7.80 (3)	7.74 (2)	7.87 (4)	7.93 (5)
	$R_1 = 5$	$R_2 = 5$	$R_3 = 9$	$R_4 = 14$	$R_5 = 12$

Source: W. G. Cochran and G. M. Cox (1957).

Since the value of L is greater than the critical value 155, we can reject H_0 at the $\alpha = .0097$ level, providing strong evidence (for the levels of potash considered) in favor of the trend of decreasing breaking strength with increasing level of potash.

For the large-sample approximation we need to compute the standardized form of L^* using (7.14). Since $k = 5$, $n = 3$, and the sample value of L is 158, we see from (7.14) that

$$L^* = \frac{158 - \left[\dfrac{3(5)(5+1)^2}{4}\right]}{\left\{\dfrac{3(5^2)(5+1)(5^2-1)}{144}\right\}^{1/2}} = \frac{158 - 135}{\sqrt{75}} = 2.66.$$

Thus, using the approximate procedure (7.15) with the value of $L^* = 2.66$ and the R command `pnorm(·)`, we see that the approximate P-value for these data is $P_0(L^* \geq 2.66) \approx 1 - \texttt{pnorm(2.66)} = 1 - .9961 = .0039$. This is in good agreement with our previous outcome using the exact test, even though n is only 3.

Comments

13. *More General Setting.* As with the Friedman procedure in Section 7.1, we could replace Assumptions A1–A3 and H_0 (7.2) with the more general null hypothesis that all possible $(k!)^n$ rank configurations for the r_{ij}'s are equally likely. Procedure (7.11) remains distribution-free for this more general hypothesis.
14. *Motivation for the Test.* If the ordering $\tau_1 < \tau_2 < \cdots < \tau_k$ is true, then R_v will tend to be larger than R_u for $u < v$. Note that L (7.10) weights R_v by the integer v and R_u by the integer u. Thus, L tends to be large when H_2 (7.9) is true, serving as partial motivation for the L test in (7.11).
15. *Assumptions.* As with the Friedman procedure in Section 7.1, we emphasize that Assumption A3 stipulates that the nk cell distributions $F_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, k$, can differ at most in their locations (medians) and that these location differences (if any) must be a result of additive block and/or treatment effects (i.e., there is no interaction between the treatment and block factors). In particular, Assumption A3 requires that the nk underlying distributions belong to the same general family (F) and that they do not differ in scale parameters (variability). We do note, however, that the test procedure (7.11) remains distribution-free under the less restrictive setting where Assumption A3 is replaced by the weaker condition A3′ stated in Comment 6. (See also Comment 13.)
16. *Special Case of Two Treatments.* For the case of $k = 2$ treatments, the procedures in (7.11) and (7.15) are equivalent to the exact and large-sample approximation forms, respectively, of the one-sided sign test, as discussed in Section 3.4.
17. *Derivation of the Distribution of L under H_0 (No-Ties Case).* The null distribution of L (7.10) can be obtained by using the fact that under H_0 (7.2), all $(k!)^n$ possible rank configurations are equally likely. As is the case for S (7.5) (see Comment 8), L does not vary with changes of the names of the blocks; however, unlike S, because it is directed toward a particular ordered alternative, the L values (in general) do change with changes of names of the treatments. Thus, building up the null distribution of L is more tedious than in the case of S. We illustrate this construction for the very special case of $k = 3$ and $n = 2$. In this

case, we need to consider $(3!)^2 = 36$ block–treatment rank configurations. We list these configurations and their associated values of $L = R_1 + 2R_2 + 3R_3$.

(a)

I	II	III
1	2	3
1	2	3

$L = 28$

(b)

I	II	III
1	2	3
1	3	2

$L = 27$

(c)

I	II	III
1	2	3
2	1	3

$L = 27$

(d)

I	II	III
1	2	3
2	3	1

$L = 25$

(e)

I	II	III
1	2	3
3	1	2

$L = 25$

(f)

I	II	III
1	2	3
3	2	1

$L = 24$

(g)

I	II	III
1	3	2
1	2	3

$L = 27$

(h)

I	II	III
1	3	2
1	3	2

$L = 26$

(i)

I	II	III
1	3	2
2	1	3

$L = 26$

(j)

I	II	III
1	3	2
2	3	1

$L = 24$

(k)

I	II	III
1	3	2
3	1	2

$L = 24$

(l)

I	II	III
1	3	2
3	2	1

$L = 23$

(m)

I	II	III
2	1	3
1	2	3

$L = 27$

(n)

I	II	III
2	1	3
1	3	2

$L = 26$

(o)

I	II	III
2	1	3
2	1	3

$L = 26$

(p)

I	II	III
2	1	3
2	3	1

$L = 24$

(q)

I	II	III
2	1	3
3	1	2

$L = 24$

(r)

I	II	III
2	1	3
3	2	1

$L = 23$

(s)

I	II	III
2	3	1
1	2	3

$L = 25$

(t)

I	II	III
2	3	1
1	3	2

$L = 24$

(u)

I	II	III
2	3	1
2	1	3

$L = 24$

(v)

I	II	III
2	3	1
2	3	1

$L = 22$

(w)

I	II	III
2	3	1
3	1	2

$L = 22$

(x)

I	II	III
2	3	1
3	2	1

$L = 21$

(y)

I	II	III
3	1	2
1	2	3

$L = 25$

(z)

I	II	III
3	1	2
1	3	2

$L = 24$

(aa)

I	II	III
3	1	2
2	1	3

$L = 24$

(bb)

I	II	III
3	1	2
2	3	1

$L = 22$

(cc)

I	II	III
3	1	2
3	1	2

$L = 22$

(dd)

I	II	III
3	1	2
3	2	1

$L = 21$

(ee)

I	II	III
3	2	1
1	2	3

$L = 24$

(ff)

I	II	III
3	2	1
1	3	2

$L = 23$

(gg)

I	II	III
3	2	1
2	1	3

$L = 23$

(hh)

I	II	III
3	2	1
2	3	1

$L = 21$

(ii)

I	II	III
3	2	1
3	1	2

$L = 21$

(jj)

I	II	III
3	2	1
3	2	1

$L = 20$

Thus, we find

$$P_0\{L=28\}=\frac{1}{36}, P_0\{L=27\}=\frac{4}{36}, P_0\{L=26\}=\frac{4}{36},$$

$$P_0\{L=25\}=\frac{4}{36}, P_0\{L=24\}=\frac{10}{36}, P_0\{L=23\}=\frac{4}{36},$$

$$P_0\{L=22\}=\frac{4}{36}, P_0\{L=21\}=\frac{4}{36}, P_0\{L=20\}=\frac{1}{36}.$$

The probability, under H_0, that L is greater than or equal to 27, for example, is therefore

$$P_0\{L\geq 27\}=P_0\{L=27\}+P_0\{L=28\}=\frac{4}{36}+\frac{1}{36}=\frac{5}{36}=.139.$$

Similarly, $P_0\{L\geq 28\}=P_0\{L=28\}=\frac{1}{36}=.028$.

Since the null distribution for L has been derived without specifying the common form (F) of the underlying distribution function for the X's under H_0 beyond the point of requiring that it be continuous, the test procedure (7.11) based on L is a distribution-free procedure. From the null distribution of L we can determine the critical value l_α and control the probability α of falsely rejecting H_0 when H_0 is true, and this error probability does not depend on the specific form of the common underlying X distribution.

For a given number of treatments k and blocks n, the R command `cPage(`α`,k,n)` can be used to find the available upper-tail critical values l_α for possible values of L. For a given available significance level α, the critical value l_α then corresponds to $P_0(L\geq l_\alpha)=\alpha$ and is given by `cPage(`α, k, n`)` $= l_\alpha$. Thus, for example, for $k=4$ and $n=5$, we have $P_0(L\geq 140)=.0106$ so that $l_{.0106}=$ `cPage(.0106, 4, 5)` $=140$ for $k=4$ and $n=5$.

18. *Calculation of the Mean and variance of L under the Null Hypothesis* H_0. In displays (7.12) and (7.13), we presented formulas for the mean and variance of L, respectively, when the null hypothesis is true. In this comment we illustrate a direct calculation of $E_0(L)$ and $\text{var}_0(L)$ in the particular case of $k=3, n=2$, and no tied observations, using the null distribution of L obtained in Comment 17. (Later, in Comment 20, we present arguments for the general derivations of $E_0(L)$ and $\text{var}_0(L)$.) The expected value of the null distribution of L is obtained directly from multiplication of each possible value of L by its probability under H_0. Thus, using the null probability values from Comment 17, we obtain

$$E_0(L)=\tfrac{1}{36}(20+28)+\tfrac{4}{36}(21+22+23+25+26+27)+\tfrac{10}{36}(24)=24.$$

This is in agreement with what we obtain using (7.12), namely,

$$E_0(L)=\frac{2(3)(3+1)^2}{4}=24.$$

A direct check on the expression for $\text{var}_0(L)$ is also easy. Again using the null probabilities from Comment 17, we have

$$\begin{aligned}\text{var}_0(L) &= E_0[\{L - E_0(L)\}^2] = E_0[\{L - 24\}^2] \\ &= \Big\{\tfrac{1}{36}[(20-24)^2 + (28-24)^2] + \tfrac{4}{36}\big[(21-24)^2 \\ &\quad + (22-24)^2 + (23-24)^2 + (25-24)^2 \\ &\quad + (26-24)^2 + (27-24)^2\big] + \tfrac{10}{36}[(24-24)^2]\Big\} \\ &= \tfrac{1}{36}(16+16) + \tfrac{4}{36}(9+4+1+1+4+9) + \tfrac{10}{36}(0) = 4,\end{aligned}$$

which agrees with what we obtain using (7.13) directly, namely,

$$\text{var}_0(L) = \frac{2(3)^2(3+1)(3^2-1)}{144} = 4.$$

19. *Exact Conditional Distribution of L under H_0 with Ties within the Blocks*. To have a test with exact significance level even in the presence of tied X's within some of the blocks, we need to consider all $(k!)^n$ possible rank configurations, where now the within-blocks ranks are obtained by using average ranks to break the ties. As in Comment 17, it still follows that under H_0 each of these $(k!)^n$ configurations is equally likely. For each such configuration, the value of L is computed and the results are tabulated. As an example, consider the case of $k = 3$, $n = 2$, and the data $X_{11} = 2.4$, $X_{12} = 3.6$, $X_{13} = 2.4$, $X_{21} = 4.0$, $X_{22} = 5.9$, and $X_{23} = 1.7$. Using average ranks to break the tie in the first block, the observed block rank vectors are $(r_{11}, r_{12}, r_{13}) = (1.5, 3, 1.5)$ and $(r_{21}, r_{22}, r_{23}) = (2, 3, 1)$. Thus, $R_1 = 3.5, R_2 = 6, R_3 = 2.5$, and the attained value of L is $3.5 + 2(6) + 3(2.5) = 23$. To assess the significance of this value of L, we would need to obtain the entire conditional null distribution of L by computing its value for each of the $(3!)^2 = 36$ equally likely (under H_0) possible configurations of the observed block rank vectors (1.5, 3, 1.5) and (2, 3, 1). This would be accomplished in exactly the same manner as is illustrated for the no-ties case in Comment 17.

20. *Large-Sample Approximation*. We can rewrite the expression for L (7.10) to obtain

$$L = \sum_{j=1}^{k} jR_j = \sum_{j=1}^{k} j\left(\sum_{i=1}^{n} r_{ij}\right) = \sum_{i=1}^{n}\left(\sum_{j=1}^{k} jr_{ij}\right) = \sum_{i=1}^{n} Q_i,$$

with $Q_i = \sum_{j=1}^{k} jr_{ij}, i = 1, \ldots, n$. Moreover, from Assumptions A1 and A3, $Q_1, \ldots, Q_n$ are mutually independent and identically distributed random variables, regardless of whether or not the null hypothesis H_0 is true. The asymptotic normality, as n tends to infinity, of the standardized form

$$L^* = \frac{L - E(L)}{[\text{var}(L)]^{1/2}} = \frac{L - nE[Q_1]}{[n\ \text{var}(Q_1)]^{1/2}} \tag{7.16}$$

then follows at once from standard central limit theory for sums of mutually independent, identically distributed random variables (cf. Randles and Wolfe (1979, p. 421)).

The computation of $E(Q_1)$ and $\text{var}(Q_1)$ is simplified by noting that

$$E(Q_1) = E\left(\sum_{j=1}^{k} j r_{1j}\right) = \sum_{j=1}^{k} jE(r_{1j}) \tag{7.17}$$

and

$$\begin{aligned}\text{var}(Q_1) &= \text{var}\left(\sum_{j=1}^{k} j r_{1j}\right) \\ &= \sum_{j=1}^{k} \text{var}(j r_{1j}) + 2\sum_{u=1}^{v-1}\sum_{v=2}^{k} \text{cov}(u r_{1u}, v r_{1v}) \\ &= \sum_{j=1}^{k} j^2 \text{var}(r_{1j}) + 2\sum_{u=1}^{v-1}\sum_{v=2}^{k} uv\ \text{cov}(r_{1u}, r_{1v}).\end{aligned} \tag{7.18}$$

In particular, when H_0 is true, $(r_{11}, \ldots, r_{1k})$ is an exchangeable random vector. Thus, under H_0, we have

$$\text{var}_0(r_{1j}) = \text{var}_0(r_{11}), \quad \text{for } j = 2, \ldots, k$$

and

$$\text{cov}_0(r_{1u}, r_{1v}) = \text{cov}_0(r_{11}, r_{12}), \quad \text{for } 1 \le u < v \le k.$$

Using these facts in (7.15) and (7.16), we obtain

$$E_0(Q_1) = E_0(r_{11}) \sum_{j=1}^{k} j = \frac{k(k+1)}{2} E_0(r_{11}) \tag{7.19}$$

and

$$\begin{aligned}\text{var}_0(Q_1) &= \text{var}_0(r_{11}) \sum_{j=1}^{k} j^2 + 2\,\text{cov}_0(r_{11}, r_{12}) \sum_{u=1}^{v-1}\sum_{v=2}^{k} uv \\ &= \frac{k(k+1)(2k+1)}{6} \text{var}_0(r_{11}) \\ &\quad + \text{cov}_0(r_{11}, r_{12}) \left[\sum_{u=1}^{k}\sum_{v=1}^{k} uv - \sum_{t=1}^{k} t^2\right] \\ &= \frac{k(k+1)(2k+1)}{6} [\text{var}_0(r_{11}) - \text{cov}_0(r_{11}, r_{12})] \\ &\quad + \left[\frac{k(k+1)}{2}\right]^2 \text{cov}_0(r_{11}, r_{12}).\end{aligned} \tag{7.20}$$

It can be shown (see Problem 22) that

$$E_0(r_{11}) = \frac{k+1}{2}, \text{var}_0(r_{11}) = \frac{k^2-1}{12} \tag{7.21}$$

and

$$\text{cov}_0(r_{11}, r_{12}) = -\frac{(k+1)}{12}. \tag{7.22}$$

Using these results in expressions (7.19) and (7.20), we obtain

$$E_0(Q_1) = \frac{k(k+1)}{2}\frac{(k+1)}{2} = \frac{k(k+1)^2}{4} \tag{7.23}$$

and

$$\text{var}_0(Q_1) = \frac{k(k+1)(2k+1)}{6}\left[\frac{k^2-1}{12} + \frac{k+1}{12}\right] + \left[\frac{k(k+1)}{2}\right]^2\left(-\frac{k+1}{12}\right),$$

which, after some straightforward algebra, yields

$$\text{var}_0(Q_1) = \frac{k^2(k+1)(k^2-1)}{144}. \tag{7.24}$$

Combining equations (7.17), (7.18), (7.23), and (7.24), we obtain

$$E_0(L) = nE_0(Q_1) = \frac{nk(k+1)^2}{4}$$

and

$$\text{var}_0(L) = \frac{nk^2(k+1)(k^2-1)}{144},$$

as stated in expressions (7.12) and (7.13), respectively. In conjunction with (7.16), this provides the justification for the approximate α level procedure in (7.15).

21. *Conservative Nature of the Large-Sample Approximation when There Are Ties within Blocks*. In applications where tied X values are observed in one or more of the blocks and average ranks are used to deal with these ties, the null variance of L based on its exact conditional null distribution (see Comment 19) is always smaller than the value obtained from expression (7.13). (This fact is illustrated in Problems 20 and 21.) As a result, the approximate level α procedure in (7.15) is conservative in the presence of within-blocks ties in the following sense: If we reject H_0 using procedure (7.15) with $\text{var}_0(L)$ obtained from expression (7.13), then we would also reject H_0 if we were to more properly use the exact conditional null variance of L in computing the value of L^* (7.14).

22. *Relation to Rank Order Correlation*. The L test is directly related to Spearman's rank order correlation coefficient r_s (8.63). Let r_i denote Spearman's correlation coefficient computed between the observed rank order and the postulated order in block i, and set $\overline{r} = \left(\sum_{i=1}^{n} r_i/n\right)$. Then, it can be shown that

$$\overline{r} = \left\{\frac{12L}{nk(k^2-1)} - \frac{3(k+1)}{(k-1)}\right\}.$$

23. *Consistency of the L Test*. Replace Assumptions A1–A3 by the less restrictive Assumption A1′ : $X_{ij} = \beta_i + e_{ij}$, where the e's are mutually independent, and Assumption A2′ : $e_{1j}, \ldots, e_{nj}$ come from the same continuous population $\prod_j, j = 1, \ldots, k$, but where $\prod_1, \ldots, \prod_k$ are not assumed to be identical. Then the test defined by (7.11) is consistent against alternatives for which $\left\{ \sum_{u<v} (v - u) p_{uv} > k(k-1)(k+1)/12 \right\}$, where $p_{uv} = P(e_{iu} < e_{iv})$ with e_{iu} a random member from $\prod_u$ and e_{iv} a random member from $\prod_v$ that is independent of e_{iu} (see Hollander (1967a)). For those situations covered by Assumptions A1–A3, this consistency statement implies the consistency statement given in Property 1.

Properties

1. *Consistency*. The test defined by (7.11) is consistent against the H_2 (7.9) alternatives. See Hollander (1967a) and Comment 23.
2. *Asymptotic Normality*. See Comment 20 and Randles and Wolfe (1979, p. 421).
3. *Efficiency*. See Hollander (1967a) and Section 7.16.

Problems

15. Brady (1969) described an experiment concerning the influence of the rhythmicity of a metronome on the speech of stutterers. The subjects were 12 severe stutterers. Each subject spoke extemporaneously for 3 min under the three conditions N, A, and R.

N: Subject spoke unaided by a metronome.

R: Subject spoke with a regular (rhythmic) metronome set at 120 ticks per minute and was instructed to pace one syllable of speech to each tick.

A: Subject spoke with an arrhythmic metronome in which the intervals between ticks ranged randomly between 0.3 and 0.7 s but with an average of 120 ticks per minute. Again the subject was instructed to pace one syllable of speech to each tick.

Table 7.6 gives the number of dysfluencies under each condition. On the basis of the conditions, and prior to looking at the data, we might expect a deviation from H_0 to be in the direction $\tau_R < \tau_A < \tau_N$. Perform Page's test using this postulated ordering.

16. Verify the relationship (see Comment 22) between L (7.10) and r_s (8.63).

Table 7.6 Influence of Rhythmicity of Metronome on Speech Fluency

	Dysfluencies under Each Condition		
Subject	*R*	*A*	*N*
1	3	5	15
2	3	3	11
3	1	3	18
4	5	4	21
5	2	2	6
6	0	2	17
7	0	2	10
8	0	3	8
9	0	2	13
10	1	0	4
11	2	4	11
12	2	1	17

Source: J. P. Brady (1969).

17. Show directly, or illustrate by means of an example, that the maximum value of L is $L_{\max} = nk(k+1)(2k+1)/6$. For what rank configuration is the maximum achieved?

18. Show directly, or illustrate by means of an example, that the minimum value of L is $L_{\min} = nk(k+1)(k+2)/6$. For what rank configuration is this minimum achieved?

19. Shelterbelts (long rows of tree plantings across the direction of the prevailing winds) have been used extensively for sometime in developed countries to protect crops and livestock from the effects of the wind. Ujah and Adeoye (1984) conducted a study to see if such shelterbelts could be used effectively to ameliorate the severe losses from droughts experienced almost annually in the arid and semiarid zones of Nigeria and considered to be a leading factor in the declining food production in Nigeria and many of its neighbors.

Ujah and Adeoye investigated the effect of shelterbelts on a variety of factors related to drought conditions, including wind velocity, air and soil temperatures, and soil moisture. The experiment was conducted at two locations about 3.5 km apart, near Dambatta. Table 7.7 presents the wind velocity data (averaged over these two locations) at various distances leeward of the shelterbelt expressed as percent wind speed reduction relative to the wind velocity on the windward side of the shelterbelt. The data are monthly (except for July, November, and December, for which the data were not available) and at leeward distances of 20, 40, 100, 150, and 200 m from the shelterbelt.

Use these data to test the hypothesis of a negative relationship between percent reduction in average wind speed and the leeward distance from a shelterbelt.

20. Consider the case of $k = 3$, $n = 2$, and the tied data set $X_{11} = 2.4$, $X_{12} = 3.6$, $X_{13} = 2.4$, $X_{21} = 4.0$, $X_{22} = 5.9$, and $X_{23} = 1.7$. What is the conditional probability distribution of L under H_0 (7.2) when average ranks are used to break within-blocks ties among the X's? (see Comment l9). How extreme is the observed value of $L = 23$ in this conditional null distribution?

21. Consider the tied data set in Problem 20 for the setting of $k = 3$ and $n = 2$. Use the conditional null probability distribution of L obtained in Problem 20 to compute the conditional null variance of L and compare this value with that of the unconditional null variance given by (7.13). Interpret these two numbers in view of the discussion in Comment 21.

22. Let $\boldsymbol{r}_1 = (r_{11}, \ldots, r_{1k})$ be a random vector of ranks that is uniformly distributed over the set of all $k!$ permutations of $(1, \ldots, k)$. Show that $E(r_{11}) = (k+1)/2$, $\text{var}(r_{11}) = (k^2-1)/12$, and $\text{cov}(r_{11}, r_{12}) = -(k+1)/12$.

23. Carry out the algebra to verify the final expression for $\text{var}_0(Q)$ in (7.24).

24. Suppose $k = 3$ and $n = 3$. Obtain the form of the exact null (H_0) distribution of L for the case of no-tied observations.

Table 7.7 Percent Reduction in Average Wind Speed at Dambatta, 1980/81

	Leeward Distance from Shelterbelt (m)				
Month	20	40	100	150	200
January	22.1	20.7	15.4	12.3	6.9
February	19.2	18.7	14.9	9.3	6.5
March	21.5	21.9	14.3	9.9	7.1
April	21.5	21.2	11.1	9.4	6.2
May	21.3	20.9	11.2	9.4	7.7
June	20.9	19.6	16.9	11.6	7.0
August	19.3	18.7	14.4	12.5	7.0
September	20.1	19.6	15.6	12.6	7.5
October	23.7	20.4	14.6	12.4	8.5

Source: J. E. Ujah and K. B. Adeoye (1984).

Table 7.8 Maximum Soil Temperature (°C) at 5-cm Depth at Dambatta, 1980/81

	Leeward Distance from Shelterbelt (m)			
Month	20	40	100	200
January	37.7	37.5	37.6	37.4
February	39.7	39.4	39.6	39.6
March	42.0	42.0	41.9	41.9
April	43.4	43.1	42.8	43.0
May	42.5	42.3	42.3	42.1
June	39.7	39.7	39.6	39.7
July	38.7	38.5	38.6	38.5
August	39.1	38.8	38.9	38.4
September	39.7	39.5	39.2	39.4
October	39.9	40.0	40.0	40.2
November	39.6	39.7	39.8	39.7

Source: J. E. Ujah and K. B. Adeoye (1984).

25. In their study of shelterbelts (see Problem 19), Ujah and Adeoye (1984) also obtained measurements of the monthly maximum soil temperature (°C) at a 5-cm depth at leeward distances of 20, 40, 100, and 200 m from the shelterbelt. These data are presented in Table 7.8.

Use these data to test the hypothesis that there is a negative relationship between maximum soil temperature at a 5-cm depth and the leeward distance from a shelterbelt.

26. For the case of $k = 2$, show that procedure (7.11) is equivalent to the exact one-sided sign test, as discussed in Section 3.4.

27. Consider the data on percentage consonants correctly identified in Table 7.4 from the study on hearing-impaired children by Nicholls and Ling (1982). From previous studies, there is reason to believe that cued speech (C) is more effective as a stand-alone method for teaching language to hearing-impaired children than lip reading (L), which, in turn, is thought to be more effective than audition (A) by itself. Find an approximate P-value using procedure (7.15) to test this conjecture.

RATIONALE FOR MULTIPLE COMPARISON PROCEDURES

In Sections 7.1 and 7.2 we have discussed procedures designed to test the null hypothesis H_0 (7.2) against either general or ordered alternatives. Upon rejection of H_0 with one of these test procedures for a given set of data, our conclusion is either that there are some unspecified differences among the treatment effects (associated with the Friedman procedure discussed in Section 7.1) or that the treatment effects follow an ordered pattern (associated with the Page procedure of Section 7.2). However, in neither of these test procedures is our conclusion pair-specific; that is, the tests in Sections 7.1 and 7.2 are not designed to enable us to reach conclusions about specific pairs of treatment effects. The relative sizes of the specific treatment effects τ_1 and τ_2, for example, cannot be inferred from the conclusions reached by either of the test procedures of Sections 7.1 or 7.2. To elicit such pairwise specific information, we turn to the class of multiple comparison procedures. In Section 7.3 we present a two-sided all-treatments multiple comparison procedure for the omnibus setting corresponding to the general alternatives H_1 (7.3). In Section 7.4 we deal with treatments-versus-control multiple comparison decisions for settings where one of the treatments plays a special role as the study control.

7.3 DISTRIBUTION-FREE TWO-SIDED ALL-TREATMENTS MULTIPLE COMPARISONS BASED ON FRIEDMAN RANK SUMS—GENERAL CONFIGURATION (WILCOXON, NEMENYI, MCDONALD-THOMPSON)

In this section we present a multiple comparison procedure based on Friedman's within-blocks ranks that is designed to make decisions about individual differences between pairs of treatment effects (τ_i, τ_j) for $i < j$, in a setting where general alternatives H_1 (7.3) are of interest. Thus, the multiple comparison procedure of this section would generally be applied to two-way layout data (with one observation per cell) *after* rejection of H_0 (7.2) with the Friedman procedure from Section 7.1. In this setting it is important to reach conclusions about all $\binom{k}{2} = k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two-sided.

Procedure

Let $R_1, \ldots, R_k$ be the treatment sums of within-blocks ranks given by (7.4). Calculate the $k(k-1)/2$ absolute differences $|R_u - R_v|, 1 \le u < v \le k$. At an experimentwise error rate of α the Wilcoxon–Nemenyi–McDonald–Thompson two-sided all-treatments multiple comparison procedure reaches its $k(k-1)/2$ pairwise decisions, corresponding to each (τ_u, τ_v) pair, $1 \le u < v \le k$, by the criterion

$$\text{Decide } \tau_u \ne \tau_v \text{ if } |R_u - R_v| \ge r_\alpha; \quad \text{otherwise decide } \tau_u = \tau_v, \tag{7.25}$$

where the constant r_α is chosen to make the experimentwise error rate equal to α; that is, r_α satisfies the restriction

$$P_0(|R_u - R_v| < r_\alpha, u = 1, \ldots, k-1; v = u+1, \ldots, k) = 1 - \alpha, \tag{7.26}$$

where the probability $P_0(.)$ is computed under H_0 (7.2). Equation (7.26) stipulates that the $k(k-1)/2$ inequalities $|R_u - R_v| < r_\alpha$ corresponding to all pairs (u, v) of treatments with $u < v$, hold simultaneously with probability $1 - \alpha$ when H_0 (7.2) is true. Comment 26 explains how to obtain the critical values r_α for k treatments, n blocks, and available experimentwise error rates α.

Large-Sample Approximation

When H_0 is true, the k-component vector $(R_1, \ldots, R_k)$ has, as n tends to infinity, an asymptotic $(k-1)$-variate normal distribution with appropriate mean vector and covariance matrix (see Comment 29 for indications of the proof). It then follows that the critical value r_α can, when the number of blocks n is large, be approximated by $[nk(k+1)/12]^{1/2} q_\alpha$, where q_α is the upper αth percentile point for the distribution of the range of k independent $N(0, 1)$ variables. Thus, the large-sample approximation for procedure (7.25) is

$$\text{Decide } \tau_u \ne \tau_v \text{ if } |R_u - R_v| \ge q_\alpha \left[\frac{nk(k+1)}{12}\right]^{1/2}; \quad \text{otherwise decide } \tau_u = \tau_v. \tag{7.27}$$

To find q_α for k treatments and a specified experimentwise error rate α, we use the R command `cRangeNor(`α, k`)`. For example, to find $q_{.05}$ for $k = 5$ treatments, we apply `cRangeNor(.05,5)` and obtain $q_{.05} = 3.858$ for $k = 5$.

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual treatment sums of ranks $R_1, \ldots, R_k$. In such cases, the experimentwise error rate associated with procedure (7.25) is only approximately equal to α.

EXAMPLE 7.3 ***Rounding First Base.***

Consider the rounding-first-base data discussed in Example 7.1. There we had found (using the large-sample approximation for the Friedman procedure) that there is strong evidence to conclude that the three methods of running to first base are not equivalent with respect to time to reach second base. To determine which of the three running methods differ in median times to second base, we apply the approximate procedure (7.27), using average ranks to break the within-runners ties in computing R_1, R_2, and R_3. Here, we have $k = 3$ and $n = 22$. For the sake of illustration, we take our approximate experimentwise error rate to be $\alpha = .01$. Using the R command `cRangeNor(`α, k`)` with $\alpha = .01$ and $k = 3$, we find `cRangeNor(.01, 3)` $= q_{.01} = 4.12$, and procedure (7.27) reduces to

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |R_u - R_v| \geq (4.12)\left[\frac{22(3)(4)}{12}\right]^{1/2} = 19.3.$$

Using the treatments sums of within-runners ranks given in Table 7.1, we find that

$$|R_2 - R_1| = 6, \quad |R_3 - R_1| = 21, \quad \text{and} \quad |R_3 - R_2| = 15.$$

Referring these absolute value rank sum differences to the approximate critical value 19.3, we see that

$$|R_2 - R_1| = 6 < 19.3 \quad \Rightarrow \quad \text{decide } \tau_2 = \tau_1,$$

$$|R_3 - R_1| = 21 \geq 19.3 \quad \Rightarrow \quad \text{decide } \tau_3 \neq \tau_1,$$

and

$$|R_3 - R_2| = 15 < 19.3 \quad \Rightarrow \quad \text{decide } \tau_3 = \tau_2.$$

Thus, at an approximate experimentwise error rate of .01, we have reached the conclusion that only the round out (treatment 1) and wide angle (treatment 3) running methods yield significantly different median times to second base. (We note that the

smallest approximate experimentwise error rate at which we would reach this conclusion is obtained by first setting

$$\max_{(u,v)} |R_u - R_v| = |R_3 - R_1| = 21 = q_\alpha \left[\frac{22(3)(4)}{12}\right]^{\frac{1}{2}}$$

and solving for

$$q_\alpha = 21 \left[\frac{22(3)(4)}{12}\right]^{-\frac{1}{2}} = 4.477.$$

Using the R command pRangeNor(q_α, k), we then find $\alpha =$ `pRangeNor(4.48, 3)` = .0044 to be the smallest experimentwise error rate at which we would decide that the round out (treatment 1) and wide angle (treatment 3) running methods yield significantly different median times to second base.

For the sake of illustration for the exact procedure in (7.25), we consider the subset of the sample data associated with the first 15 baseball players in Table 7.1. For that subset we have $k = 3$, $n = 15$, and the three treatment sums of ranks $R_1^* = 37$, $R_2^* = 31$, and $R_3^* = 22$. With $k = 3$, $n = 15$, and experimentwise error rate $\alpha = .047$, we apply the R command cWNMT(α,k,n) and find `cWNMT(.047, 3, 15)` = 13. Thus, we have $r_{.047} = 13$ and procedure (7.25) becomes

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |R_u^* - R_v^*| \geq 13.$$

Since $|R_2^* - R_1^*| = 6, |R_3^* - R_1^*| = 15$, and $|R_3^* - R_2^*| = 9$, we see that our decisions for this subset of data using procedure (7.25) would be $\tau_2 = \tau_1, \tau_3 \neq \tau_1$, and $\tau_3 = \tau_2$, in agreement with what we found using the entire set of 22 baseball players and the approximate procedure (7.27). (Note, however, that for this smaller set of data, we could no longer conclude that $\tau_3 \neq \tau_1$ at an experimentwise error rate as low as .01.)

Comments

24. *Rationale for Multiple Comparison Procedures*. We think of the methods of this section as multiple comparison procedures. The aim of applying such procedures goes beyond the point of deciding whether the treatments are equivalent to the (often more important) problem of selecting which, if any, treatments differ from one another. Thus, the user makes $k(k-1)/2$ decisions, one for each pair of treatments. Equation (7.26) states that the probability of making all correct decisions when H_0 is true is controlled to be $1 - \alpha$; that is, when using procedure (7.25), the probability of at least one incorrect decision, when H_0 is true, is controlled to be α. This error rate is derived under the assumption that H_0 is true, but it does not depend on the particular underlying distributional form F. This is why we call (7.25) a distribution-free multiple comparison procedure.

    The multiple comparison procedures of this section can also be interpreted as hypothesis tests. If we consider the procedure that rejects H_0 if and only if the inequality of (7.25) [or of (7.27)] holds for at least one (u, v) pair, $1 \leq u < v \leq k$, this is a distribution-free test of size α for H_0 (7.2).

25. *Experimentwise Error Rate*. The use of an experimentwise error rate represents a very conservative approach to multiple comparisons. We are insisting that the probability of making only correct decisions be $1-\alpha$ when the null hypothesis H_0 (7.2) of treatment equivalence is true. Thus, although we have a high degree of protection when H_0 is true, we often apply such techniques where we have evidence (perhaps based on a priori information or perhaps obtained by applying the Friedman test, as in Example 7.3) that H_0 is not true. This protection under H_0 also makes it harder for the procedure to judge treatments as differing significantly when in fact H_0 is false, and this difficulty becomes more severe as k increases. See Comment 6.54 for additional discussion of experimentwise error rates.

26. *Critical Values* r_α. The r_α critical values can be obtained by using the fact that under H_0 (7.2), all $(k!)^n$ rank configurations are equally likely. Thus, to obtain the probability under H_0 that $|R_u - R_v| < c$ simultaneously for $u = 1, \ldots, k-1$ and $v = u+1, \ldots, k$, we can count the number of configurations for which the event $B = \{|R_u - R_v| < c, u = 1, \ldots, k-1; v = u+1, \ldots, k\}$ occurs and divide this number by $(k!)^n$. For an illustration, consider the 24 configurations of Comment 8, corresponding to the case $k = 4$, $n = 2$. (As in Comment 8, the same reasoning enables us to consider only 24 rather than $(4!)^2 = 576$ configurations.) For each configuration, we now display the values of $|R_1 - R_2|, |R_1 - R_3|, |R_1 - R_4|, |R_2 - R_3|, |R_2 - R_4|$, and $|R_3 - R_4|$.

(a)	$\|R_1-R_2\|=2$	(b)	$\|R_1-R_2\|=2$	(c)	$\|R_1-R_2\|=3$	(d)	$\|R_1-R_2\|=3$
	$\|R_1-R_3\|=4$		$\|R_1-R_3\|=5$		$\|R_1-R_3\|=5$		$\|R_1-R_3\|=3$
	$\|R_1-R_4\|=6$		$\|R_1-R_4\|=5$		$\|R_1-R_4\|=4$		$\|R_1-R_4\|=6$
	$\|R_2-R_3\|=2$		$\|R_2-R_3\|=3$		$\|R_2-R_3\|=2$		$\|R_2-R_3\|=0$
	$\|R_2-R_4\|=4$		$\|R_2-R_4\|=3$		$\|R_2-R_4\|=1$		$\|R_2-R_4\|=3$
	$\|R_3-R_4\|=2$		$\|R_3-R_4\|=0$		$\|R_3-R_4\|=1$		$\|R_3-R_4\|=3$
(e)	$\|R_1-R_2\|=4$	(f)	$\|R_1-R_2\|=4$	(g)	$\|R_1-R_2\|=0$	(h)	$\|R_1-R_2\|=0$
	$\|R_1-R_3\|=3$		$\|R_1-R_3\|=4$		$\|R_1-R_3\|=3$		$\|R_1-R_3\|=4$
	$\|R_1-R_4\|=5$		$\|R_1-R_4\|=4$		$\|R_1-R_4\|=5$		$\|R_1-R_4\|=4$
	$\|R_2-R_3\|=1$		$\|R_2-R_3\|=0$		$\|R_2-R_3\|=3$		$\|R_2-R_3\|=4$
	$\|R_2-R_4\|=1$		$\|R_2-R_4\|=0$		$\|R_2-R_4\|=5$		$\|R_2-R_4\|=4$
	$\|R_3-R_4\|=2$		$\|R_3-R_4\|=0$		$\|R_3-R_4\|=2$		$\|R_3-R_4\|=0$
(i)	$\|R_1-R_2\|=2$	(j)	$\|R_1-R_2\|=2$	(k)	$\|R_1-R_2\|=3$	(l)	$\|R_1-R_2\|=3$
	$\|R_1-R_3\|=4$		$\|R_1-R_3\|=1$		$\|R_1-R_3\|=1$		$\|R_1-R_3\|=3$
	$\|R_1-R_4\|=2$		$\|R_1-R_4\|=5$		$\|R_1-R_4\|=4$		$\|R_1-R_4\|=2$
	$\|R_2-R_3\|=2$		$\|R_2-R_3\|=1$		$\|R_2-R_3\|=2$		$\|R_2-R_3\|=0$
	$\|R_2-R_4\|=0$		$\|R_2-R_4\|=3$		$\|R_2-R_4\|=1$		$\|R_2-R_4\|=1$
	$\|R_3-R_4\|=2$		$\|R_3-R_4\|=4$		$\|R_3-R_4\|=3$		$\|R_3-R_4\|=1$
(m)	$\|R_1-R_2\|=1$	(n)	$\|R_1-R_2\|=1$	(o)	$\|R_1-R_2\|=0$	(p)	$\|R_1-R_2\|=0$
	$\|R_1-R_3\|=1$		$\|R_1-R_3\|=3$		$\|R_1-R_3\|=3$		$\|R_1-R_3\|=0$
	$\|R_1-R_4\|=4$		$\|R_1-R_4\|=2$		$\|R_1-R_4\|=1$		$\|R_1-R_4\|=4$
	$\|R_2-R_3\|=2$		$\|R_2-R_3\|=4$		$\|R_2-R_3\|=3$		$\|R_2-R_3\|=0$
	$\|R_2-R_4\|=5$		$\|R_2-R_4\|=3$		$\|R_2-R_4\|=1$		$\|R_2-R_4\|=4$
	$\|R_3-R_4\|=3$		$\|R_3-R_4\|=1$		$\|R_3-R_4\|=2$		$\|R_3-R_4\|=4$

(q)	$\lvert R_1 - R_2\rvert = 2$	(r)	$\lvert R_1 - R_2\rvert = 2$	(s)	$\lvert R_1 - R_2\rvert = 2$	(t)	$\lvert R_1 - R_2\rvert = 2$
	$\lvert R_1 - R_3\rvert = 0$		$\lvert R_1 - R_3\rvert = 1$		$\lvert R_1 - R_3\rvert = 0$		$\lvert R_1 - R_3\rvert = 1$
	$\lvert R_1 - R_4\rvert = 2$		$\lvert R_1 - R_4\rvert = 1$		$\lvert R_1 - R_4\rvert = 2$		$\lvert R_1 - R_4\rvert = 1$
	$\lvert R_2 - R_3\rvert = 2$		$\lvert R_2 - R_3\rvert = 1$		$\lvert R_2 - R_3\rvert = 2$		$\lvert R_2 - R_3\rvert = 3$
	$\lvert R_2 - R_4\rvert = 0$		$\lvert R_2 - R_4\rvert = 1$		$\lvert R_2 - R_4\rvert = 4$		$\lvert R_2 - R_4\rvert = 3$
	$\lvert R_3 - R_4\rvert = 2$		$\lvert R_3 - R_4\rvert = 0$		$\lvert R_3 - R_4\rvert = 2$		$\lvert R_3 - R_4\rvert = 0$
(u)	$\lvert R_1 - R_2\rvert = 1$	(v)	$\lvert R_1 - R_2\rvert = 1$	(w)	$\lvert R_1 - R_2\rvert = 0$	(x)	$\lvert R_1 - R_2\rvert = 0$
	$\lvert R_1 - R_3\rvert = 1$		$\lvert R_1 - R_3\rvert = 1$		$\lvert R_1 - R_3\rvert = 1$		$\lvert R_1 - R_3\rvert = 0$
	$\lvert R_1 - R_4\rvert = 2$		$\lvert R_1 - R_4\rvert = 0$		$\lvert R_1 - R_4\rvert = 1$		$\lvert R_1 - R_4\rvert = 0$
	$\lvert R_2 - R_3\rvert = 0$		$\lvert R_2 - R_3\rvert = 2$		$\lvert R_2 - R_3\rvert = 1$		$\lvert R_2 - R_3\rvert = 0$
	$\lvert R_2 - R_4\rvert = 3$		$\lvert R_2 - R_4\rvert = 1$		$\lvert R_2 - R_4\rvert = 1$		$\lvert R_2 - R_4\rvert = 0$
	$\lvert R_3 - R_4\rvert = 3$		$\lvert R_3 - R_4\rvert = 1$		$\lvert R_3 - R_4\rvert = 2$		$\lvert R_3 - R_4\rvert = 0$

Thus, for example,

$$\begin{aligned} P_0\{|R_u - R_v| < 6, u = 1,2,3; v = u+1,\ldots,4\} &= P_0\{|R_1 - R_2| < 6; |R_1 - R_3| < 6; |R_1 - R_4| < 6; \\ &\quad |R_2 - R_3| < 6; |R_2 - R_4| < 6; |R_3 - R_4| < 6\} \\ &= \tfrac{22}{24} = 1 - .083, \end{aligned}$$

because for 22 of the configurations—all but configurations (a) and (d)—the event $\{|R_1 - R_2| < 6; |R_1 - R_3| < 6; |R_1 - R_4| < 6; |R_2 - R_3| < 6; |R_2 - R_4| < 6; |R_3 - R_4| < 6\}$ occurs. This .083 probability agrees with the result obtained from using the R command `cWNMT(`α`,k,n)`; that is, `cWNMT(.083, 4, 2)` $= r_{.083} = 6$ for $k = 4$ treatments and $n = 3$ blocks.

27. *Relationship to Range of Rank Sums*. Define the range of $R_1, \ldots, R_k$ as

$$\text{Range } [R_1, \ldots, R_k] = \max[R_1, \ldots, R_k] - \min[R_1, \ldots, R_k].$$

Then, $|R_u - R_v|$ is less than c, for all $u < v$, if and only if range $[R_1, \ldots, R_k]$ is less than c. Thus, in Comment 26, instead of computing the values of $|R_1 - R_2|, |R_1 - R_3|, |R_1 - R_4|, |R_2 - R_3|, |R_2 - R_4|, |R_3 - R_4|$ for each rank configuration, we need to have calculated only range $[R_1, \ldots, R_k]$ for each configuration. That is, we can obtain the critical constants r_α by computing only the range for each possible rank configuration. We do, however, need to compute the individual absolute differences $|R_u - R_v|$ in order to apply procedure (7.25) to a set of data.

28. *Historical Development*. The basic idea behind the multiple comparison procedures (7.25) and (7.27) based on the Friedman rank sums is attributed by McDonald and Thompson (1967) to Wilcoxon who "... in 1956, in an unpublished notebook, carried out the first correct probability computation for 3 objects (treatments) and 3 judges (blocks)..." Nemenyi (1963) obtained a small number of exact critical values r_α for procedure (7.25) in his Ph.D. dissertation. McDonald and Thompson (1967) provided additional exact critical values.

29. *Large-Sample Approximation*. Let $\boldsymbol{R} = (R_1, \ldots, R_k)$ be the vector of the Friedman rank sums. Then, it can be shown that a properly standardized $\boldsymbol{R}$ has, as n tends to infinity, an asymptotic multivariate normal distribution with appropriate mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (see Miller (1966) for details). It follows directly from this result (again, see Miller (1966)) that the procedure in (7.27) has an asymptotic experimentwise error rate equal to α.
30. *Dependence on Observations from Other Noninvolved Treatments*. The absolute difference $|R_u - R_v|$ depends on the values of the observations from the other $k - 2$ treatments, in addition to the observations from treatments u and v. Thus, the multiple comparison procedures (7.25) and (7.27) both have the disadvantage that the decision concerning treatment u and treatment v can be affected by changes only in the observations from one or more of the other $k - 2$ treatments that are not directly involved. This difficulty has been emphasized by Miller (1966) and Gabriel (1969).
31. *Approximately Distribution-Free Multiple Comparison Procedures Based on Signed Ranks*. The multiple comparison procedures in (7.25) and (7.27) both utilize the within-blocks Friedman ranking schemes, and, as a result, they are related to the sign procedures for paired replicates data (see Comments 7 and 16). Competitor all-treatments multiple comparison procedures based on signed ranks that utilize between-block information are discussed in Section 7.13. These signed rank procedures are, however, only asymptotically distribution-free.

Properties

1. *Asymptotic Multivariate Normality*. See Miller (1966).
2. *Efficiency*. See Section 7.16.

Problems

28. Livesey (1967) has compared the performance of rats, Rabbits, and cats on the Hebb–Williams (1946) elevated pathway test (EPT). Table 7.9, based on a subset of the Livesey data, gives mean error scores by species for 12 problems. Using procedure (7.25), find the species (if any) that differ significantly.

29. For the case of $k = 3$, $n = 9$, and $\alpha = .05$, compare procedures (7.25) and (7.27).

30. Consider the rounding-first-base data discussed in Examples 7.1 and 7.3. Using the data for all 22 players, find the smallest approximate experimentwise error rate at which we would declare that the median time to second base for the narrow angle method of running is different from that for the wide angle method of running.

31. Apply procedure (7.25) to the ozone exposure data of Table 7.2.

32. Apply the approximate procedure (7.27) to the percentage of correctly identified consonants data of Table 7.4.

33. Find the totality of all available experimentwise error rates α and the associated critical values r_α for procedure (7.25) when $k = 3$ and $n = 3$.

34. Illustrate the difficulty discussed in Comment 30 by means of a numerical example.

35. Consider the serum CPK data of Table 7.3 in Problem 5. Find the smallest (available) experimentwise error rate at which the most significant difference in treatment effects between the time measurements would be detected.

Table 7.9 Error Scores by Species

Problem Number	Rats	Rabbits	Cats
1	1.5	1.7	0.3
2	1.1	1.5	1.0
3	1.8	8.1	3.6
4	1.9	1.3	0.0
5	4.3	4.0	0.6
6	2.0	4.6	5.5
7	8.4	4.0	1.0
8	6.6	5.1	3.1
9	2.4	2.5	0.1
10	6.5	6.9	1.6
11	2.6	2.5	4.3
12	6.5	6.8	1.0

Source: P. J. Livesey (1967).

36. Consider the serum CPK data of Table 7.3 in Problem 5. Find the smallest (available) experimentwise error rate at which we would declare that the typical serum CPK activity at 19 h postexercise is different from that at 42 h postexercise.

7.4 DISTRIBUTION-FREE ONE-SIDED TREATMENTS VERSUS CONTROL MULTIPLE COMPARISONS BASED ON FRIEDMAN RANK SUMS (NEMENYI, WILCOXON–WILCOX, MILLER)

In this section we turn our attention to a multiple comparison procedure designed to make decisions about individual differences between the median effect for a single, baseline control population and the median effects of each of the remaining $k-1$ treatments. This treatments-versus-control multiple comparison procedure can be applied *after* rejection of H_0 (7.2) with either the Friedman or the Page test discussed in Sections 7.1 and 7.2, respectively. Its application leads to conclusions about the differences between each of the $k-1$ treatment effects and the control effect, and these conclusions are naturally one-sided.

Procedure

For simplicity of notation, we let treatment 1 assume the role of the single baseline control. Let $R_1, \ldots, R_k$ be the treatment sums of the within-blocks ranks given by (7.4). Calculate the $k-1$ differences $(R_u - R_1), u = 2, \ldots, k$. At an experimentwise error rate of α, the Nemenyi–Wilcoxon–Wilcox–Miller one-sided treatments-versus-control multiple comparison procedure reaches its $k-1$ pairwise decisions, corresponding to each (τ_1, τ_u) pair, for $u = 2, \ldots, k$, by the criterion

$$\text{Decide } \tau_u > \tau_1 \text{ if } (R_u - R_1) \geq r_\alpha^*; \text{ otherwise decide } \tau_u = \tau_1, \tag{7.28}$$

where the constant r_α^* is chosen to make the experimentwise error rate equal to α; that is, r_α^* satisfies the restriction

$$P_0((R_u - R_1) < r_\alpha^*, u = 2, \ldots, k) = 1 - \alpha, \tag{7.29}$$

where the probability $P_0(.)$ is computed under H_0 (7.2). Equation (7.29) stipulates that the $k-1$ inequalities $(R_u - R_1) < r_\alpha^*$, corresponding to each treatment paired with the control, hold simultaneously with probability $1-\alpha$ when H_0(7.2) is true. Comment 35 explains how to obtain the critical value r_α^* for k treatments, n blocks, and available experimentwise error rates α. (For discussion of how to adjust procedure (7.28) for settings where it is of interest to decide whether the treatment effects are *smaller* than the control effect, see Comment 34.)

Large-Sample Approximation

When H_0 is true, the $(k-1)$-component vector $(R_2 - R_1, \ldots, R_k - R_1)$ has, as n tends to infinity, an asymptotic $(k-1)$-variate normal distribution with mean vector $\mathbf{0}$ (see Comment 37 for an indication of the proof). It then follows that the critical value r_α^* can, when the number of blocks is large, be approximated by $[nk(k+1)/6]^{1/2} m_{\alpha,1/2}^*$, where $m_{\alpha,1/2}^*$ is the upper αth percentile point for the distribution of the maximum of $(k-1)$ $N(0,1)$ random variables with common correlation $\rho = \frac{1}{2}$. Thus, the large-sample approximation for procedure (7.28) is

$$\begin{aligned} &\text{Decide } \tau_u > \tau_1 \text{ if } (R_u - R_1) \geq [nk(k+1)/6]^{1/2} m_{\alpha,1/2}^*; \\ &\text{otherwise decide } \tau_u = \tau_1. \end{aligned} \tag{7.30}$$

To find $m_{\alpha,1/2}^*$ for k treatments and a specified experimentwise error rate α, we use the R command `cMaxCorrNor(`α`,k,0.5)`. For example, to find $m_{.04584,1/2}^*$ for $k=4$ treatments, we apply `cMaxCorrNor(.04584,4,0.5)` and obtain $m_{.04584,1/2}^* = 2.19$.

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual treatment sums of ranks $R_1, \ldots, R_k$. In such cases, the experimentwise error rate associated with procedure (7.28) is only approximately equal to α.

EXAMPLE 7.4 ***Stuttering Adaptation.***

Daly and Cooper (1967) considered the rate of stuttering adaptation under three conditions. Eighteen subjects (college-age stutterers) read each of three different passages five consecutive times. In one condition, electroshock was administered during each moment of stuttering, and in another condition, electroshock was administered immediately following each stuttered word. The remaining condition was a control with no electroshock. The percentage of stuttering behavior during each reading was recorded, and Table 7.10 presents for each subject a rate of adaptation score under each condition. The score was found by using the residual measurement method suggested by Tate, Cullinan, and Ahlstrand (1961).

To determine if either of the two treatments yield improved (larger) median adaptation scores, we apply procedure (7.28), using average ranks to break the within-subjects ties in computing R_1, R_2, and R_3. Here, we have $k=3$ and $n=18$. For the sake of illustration, we take our experimentwise error rate to be $\alpha = .0492$. Using the R command

Table 7.10 Adaptation Scores for College-Age Stutterers

	Treatment		
Subject	1 (No shock)	2 (Shock following)	3 (Shock during)
1	57 (3)	38 (1)	51 (2)
2	59 (3)	48 (1)	56 (2)
3	44 (1.5)	50 (3)	44 (1.5)
4	51 (2)	53 (3)	44 (1)
5	43 (1)	53 (3)	50 (2)
6	49 (1)	56 (3)	54 (2)
7	48 (2)	37 (1)	50 (3)
8	56 (2)	58 (3)	40 (1)
9	44 (1.5)	44 (1.5)	50 (3)
10	50 (2)	50 (2)	50 (2)
11	44 (1)	58 (3)	56 (2)
12	50 (3)	48 (2)	46 (1)
13	70 (2)	60 (1)	74 (3)
14	42 (1)	58 (3)	57 (2)
15	58 (1)	60 (2)	74 (3)
16	54 (3)	38 (1)	48 (2)
17	38 (1)	48 (2.5)	48 (2.5)
18	48 (2)	56 (3)	44 (1)
	$R_1 = 33$	$R_2 = 39$	$R_3 = 36$

Source: D. A. Daly and E. B. Cooper (1967).

`cNWWM`(α,k,n) with $k = 3$ and $n = 18$, we find `cNWWM(.0492,3,18)` $= r^*_{.0492} = 12$, and procedure (7.28) reduces to

$$\text{Decide } \tau_u > \tau_1 \text{ if } (R_u - R_1) \geq 12.$$

Using the treatments sums of within-subjects ranks given in Table 7.10, we find that

$$(R_2 - R_1) = 6 \quad \text{and} \quad (R_3 - R_1) = 3.$$

Referring these rank sum differences to the critical value 12, we see that

$$(R_2 - R_1) = 6 < 12 \quad \Rightarrow \quad \text{decide } \tau_2 = \tau_1,$$
$$(R_3 - R_1) = 3 < 12 \quad \Rightarrow \quad \text{decide } \tau_3 = \tau_1.$$

Thus, at an experimentwise error rate of .0492, we find no statistical evidence that either of the two electroshock treatments lead to an increase in median adaptation scores over the control setting. (In fact, we can use the `R` command `pNWWM`$(\cdot)$ to obtain the smallest experimentwise error rate at which we would be able to declare a statistically significant increase in median adaptation scores for either of the two treatments. Since the largest observed difference in rank sums is $(R_2 - R_1) = 6$, we see that the smallest experimentwise error rate at which we could declare a statistically significant increase in median adaptation scores for either of the two treatments is `pNWWM(adaptation.scores)` `$p.val[1]` = .2859.

For the sake of illustration for the large-sample approximation (7.30), we simply note that $m^*_{.02002,1/2} =$ `cMaxCorrNor(.02002,3,0.5)` $= 2.30$ for $k = 3$ and $m^*_{.05410,1/2} =$ `cMaxCorrNor(.05410,6,0.5)` $= 2.20$ for $k = 6$.

Comments

32. *Rationale for Treatments-versus-Control Multiple Comparison Procedures*. The general rationale for the multiple comparison procedures of this section is the same as that given in Comment 24 for the two-sided all-treatments multiple comparison procedures of Section 7.3. The only additional factor here is that the treatments-versus-control procedures of this section do not compare all treatments, but only each noncontrol treatment with the control on a directional basis. This situation arises, for example, in drug screening in the examination of many new treatments in hopes of improving on a standard, and there is no initial reason to perform between treatment comparisons. Of course, similar comparisons between treatments that were selected as being better than the control would most likely be carried out later in a follow-up study.

    The multiple comparison procedures of this section, which involve making $k-1$ decisions, can also be interpreted as hypothesis tests. If we consider the procedure that rejects H_0 if and only if the inequality in (7.28) [or in (7.30)] holds for at least one $(1, u)$ pair, $u = 2, \ldots, k$, then this is a distribution-free test of level α for H_0 (7.2).

33. *Experimentwise Error Rate*. The use of an experimentwise error rate represents a very conservative approach to multiple comparisons. We insist that the probability of making only correct decisions be $1-\alpha$ when the hypothesis H_0 (7.2) of treatment equivalence is true. Thus we have a high degree of protection when H_0 is true, but we often apply the techniques of this section when we have evidence (perhaps based on a priori information or perhaps obtained by applying a previous test procedure) that H_0 is not true. (For additional general remarks about experimentwise error rates, see Comment 6.54.)

34. *Opposite Direction Decisions*. Procedures (7.28) and (7.30) are designed for the one-sided case where the decisions are $\tau_u > \tau_1$ versus $\tau_u = \tau_1, u = 2, \ldots, k$. To handle the analogous one-sided situation where the decisions involve $\tau_u < \tau_1$ versus $\tau_u = \tau_1, u = 2, \ldots, k$, use (7.28) and (7.30) with $(R_u - R_1)$ replaced by $(R_1 - R_u)$ for $u = 2, \ldots, k$.

35. *Critical Values r_α^**. The r_α^* critical values can be obtained by using the fact that under H_0 (7.2), all $(k!)^n$ rank configurations are equally likely. However, the computational effort is greater in this treatments-versus-control problem than in the all-treatments problem, because the values $R_u - R_1, u = 2, \ldots, k$, are in general changed when we relabel the control treatment. (In the all-treatments case, the relevant statistic range $[R_1, \ldots, R_k]$ is unaffected by treatment relabelings. (See Comment 27.)

    Let us now do an example to illustrate the nature of the necessary computations. For simplicity, we take the case $n = 3, k = 3$. Here, the largest possible value of $R_3 - R_1$ is 6, corresponding to the configuration

(a)	I	II	III
	1	2	3
	1	2	3
	1	2	3,

where $R_1 = 3$, $R_3 = 9$. Similarly, the largest possible value of $R_2 - R_1$ is 6, corresponding to

(b)	I	II	III
	1	3	2
	1	3	2
	1	3	2.

Since none of the other configurations can yield an $R_u - R_1$ difference as large as 6, we have $P_0\{(R_2 - R_1) \geq 6 \text{ or } (R_3 - R_1) \geq 6\} = 2/[(3!)^3] = 2/216 = .0093$. Thus, in the notation of (7.28), we have $r^*_{.0093} = 6$, in agreement with the result obtained from using the R command `cNWWM(`α`,`k`,`n`)`; that is, `cNWWM(.0093,3,3)` $= r^*_{.0093} = 6$ for $k = 3$ treatments and $n = 3$ blocks.

36. *Historical Development*. The basic idea behind the treatments-versus-control multiple comparison procedures (7.28) and (7.30) based on the Friedman rank sums was initially discussed by Nemenyi (1963), Wilcoxon and Wilcox (1964), and Miller (1966). Windham (1971) provided the exact critical values r^*_α for procedure (7.28) for the case of $k = 3$, $n = 2(1)18$ and for $k = 4$, $n = 2(1)5$. Additional values of r^*_α were obtained by Odeh (1977) for the settings $k = 2(1)5$, $n = 2(1)8$ and $k = 6$, $n = 2(1)6$.

37. *Large-Sample Approximation.* Let $\boldsymbol{R}_d = (R_2 - R_1, \ldots, R_k - R_1)$ be the vector of differences between the treatment rank sums and the control rank sum. Then, it can be shown that a properly standardized $\boldsymbol{R}_d$ has, as n tends to infinity, an asymptotic multivariate normal distribution with mean vector $\mathbf{0}$ and appropriate covariance matrix Σ (see Miller (1966) for details). It follows directly from this result (again, see Miller (1966)) that the procedure in (7.30) has an asymptotic experimentwise error rate equal to α.

38. *Dependence on Observations from Other Noninvolved Treatments*. The treatments-versus-control multiple comparison procedures of this section suffer from the same disadvantage mentioned in Comment 30 for the corresponding all-treatments multiple comparisons. The decision between treatment $u(>1)$ and the control can be affected by changes only in the observations from one or more of the other $k - 2$ treatments that are not directly involved.

39. *Two-Sided Treatments-versus-Control Multiple Comparison Procedures*. The multiple comparison procedures of this section are both one sided by nature, resulting in decisions between $\tau_u = \tau_1$ and $\tau_u > \tau_1$ for every $u = 2, \ldots, k$ (or between $\tau_u = \tau_1$ and $\tau_u < \tau_1$ for every $u = 2, \ldots, k$, as noted in Comment 34). We view such one-sided comparisons to be the most natural approach for treatments-versus-control settings. In such situations, we are generally interested in seeing which, if any, of the proposed new treatments are better than a standard control or placebo. In most practical applications, *better* is synonymous with one-sided comparisons (all in one direction or all in the other), and thus our emphasis on such procedures in this section. However, a two-sided treatments-versus-control analog to procedure (7.28) has been developed in the literature and corresponds to the criterion

$$\text{Decide } \tau_u \neq \tau_1 \text{ if } |R_u - R_1| \geq r^{**}_\alpha; \quad \text{otherwise decide } \tau_u = \tau_1, \tag{7.31}$$

where the constant r_α^{**} is chosen to make the experimentwise error rate equal to α; that is,

$$P_0\{|R_u - R_1| < r_\alpha^{**}, u = 2, \ldots, k\} = 1 - \alpha,$$

where the probability $P_0(.)$ is computed under H_0 (7.2). Windham (1971) provided values of r_α^{**} for procedure (7.31) for the settings $k = 3$, $n = 2(1)18$ and $k = 4$, $n = 2(1)5$ (see also Hollander and Wolfe (1973)). A large-sample approximation to (7.31) is discussed in Miller (1966).

40. *Approximately Distribution-Free Multiple Comparison Procedures Based on Signed Ranks*. The multiple comparison procedures (7.28) and (7.30) both utilize the within-blocks Friedman ranking schemes, and, as a result, they are related to the sign procedures for paired replicates data (see Comments 7 and 16). Competitor treatments-versus-control multiple comparison procedures based on signed ranks that utilize between-block information are discussed in Section 7.14. These signed rank procedures are, however, only asymptotically distribution-free.

Properties

1. *Asymptotic Multivariate Normality*. See Miller (1966).
2. *Efficiency*. See Section 7.16.

Problems

37. Consider the serum CPK activity data from Problem 5. Treating preexercise as a control and ignoring the peak psychotic period data, apply procedure (7.28) to decide if there is statistical evidence of increased serum CPK activity either 19 or 42 h after exercise.

38. Apply an appropriate one-sided multiple comparison procedure (see (7.28) and Comment 34) to the rhythmicity data of Table 7.6, letting the condition N (subject spoke unaided by a metronome) serve as the control.

39. For the case $k = 3$, $n = 18$, and $\alpha \approx .01$, compare procedures (7.28) and (7.30).

40. Consider the rounding-first-base data discussed in Examples 7.1 and 7.3. Using the data for all 22 players and treating the round out method of running to second base as the control, find the smallest approximate experimentwise error rate at which we would declare that the median time to second base for the wide angle method of running is smaller than that for the round out method.

41. Illustrate the difficulty discussed in Comment 38 by means of a numerical example.

42. Find the complete list of available experimentwise error rates α and the associated r_α^* critical values for procedure (7.28) when $k = 3$ and, $n = 2$.

43. Consider the subset of the data on percentage correctly identified consonants in Table 7.4 corresponding to conditions A, AL, and AC. Treating condition A as a control, find the smallest experimentwise error rate for procedure (7.28) at which we would detect the condition (L or C) yielding the most improvement in performance when added to A in the syllable presentation.

44. Treating condition A as a control, apply procedure (7.30) to the data on percentage correctly identified consonants in Table 7.4.

45. Consider the maximum soil temperature data in Table 7.8. Apply the appropriate treatments-versus-control procedure to decide if maximum soil temperature is significantly warmer at 20, 40, or 100 m from the shelterbelt than at a distance of 200 m.

7.5 CONTRAST ESTIMATION BASED ON ONE-SAMPLE MEDIAN ESTIMATORS (DOKSUM)

In this section we discuss a method for point estimation of certain linear combinations of treatment effects known in the literature as *contrasts*. We define such a contrast in the treatment effects $\tau_1, \ldots, \tau_k$ to be any linear combination of the form

$$\theta = \sum_{i=1}^{k} a_i \tau_i, \tag{7.32}$$

where $a_1, \ldots, a_k$ are any specified constants such that $\sum_{i=1}^{k} a_i = 0$. Equivalently, we can write θ in terms of the individual differences in treatment effects (known in the literature as *simple contrasts*)

$$\Delta_{hj} = \tau_h - \tau_j, \quad h = 1, \ldots, k; \quad j = 1, \ldots, k, \tag{7.33}$$

by noting that

$$\theta = \sum_{h=1}^{k} \sum_{j=1}^{k} d_{hj} \Delta_{hj}, \tag{7.34}$$

where

$$d_{hj} = \frac{a_h}{k}, \quad h = 1, \ldots, k; \quad j = 1, \ldots, k. \tag{7.35}$$

For a given setting, decisions about which contrasts to estimate can be related to either a priori interest in particular linear combinations of the τ's or the results of one of the multiple comparison procedures discussed in Sections 7.3 and 7.4.

Procedure

For each pair of treatments $(u, v), u \neq v = 1, \ldots, k$, compute the differences

$$D_{uv}^{i} = X_{iu} - X_{iv}, i = 1, \ldots, n, \tag{7.36}$$

between the treatment u and treatment v observations for each of the n blocks. For $1 \leq u \neq v \leq k$, let

$$Z_{uv} = \text{median } \{D_{uv}^{i}, i = 1, \ldots, n\}. \tag{7.37}$$

Since $Z_{vu} = -Z_{uv}$, we need only to calculate the $k(k-1)/2$ values Z_{uv} corresponding to $u < v$. We refer to Z_{uv} as the "unadjusted" estimator of the simple contrast $\Delta_{uv} = \tau_u - \tau_v$. (Note that Z_{uv} is just the median estimator of Section 3.5, applied here to the $X_{iu} - X_{iv}$ differences. For example, Z_{23} is the median of the $X_{i2} - X_{i3}$ differences, $i = 1, \ldots, n$, and is the "unadjusted" estimator of the simple contrast $\tau_2 - \tau_3$.) Next, we compute

$$Z_{u.} = \sum_{j=1}^{k} \frac{Z_{uj}}{k}, \quad u = 1, \ldots, k, \tag{7.38}$$

where we note that $Z_{uu} = 0$ for $u = 1, \ldots, k$. Setting

$$\tilde{\Delta}_{uv} = Z_{u.} - Z_{v.}, \tag{7.39}$$

the adjusted estimator of θ is given by

$$\tilde{\theta} = \sum_{j=1}^{k} a_j Z_{j.}, \tag{7.40}$$

or, equivalently,

$$\tilde{\theta} = \sum_{h=1}^{k} \sum_{j=1}^{k} d_{hj} \tilde{\Delta}_{hj}. \tag{7.41}$$

EXAMPLE 7.5 ***Rounding First Base.***

Consider the rounding-first-base data originally presented in Table 7.1 of Example 7.1. We illustrate the Doksum contrast estimator $\tilde{\theta}$ (7.40) on the simple contrast $\theta = \tau_{\text{roundout}} - \tau_{\text{wide angle}} = \tau_1 - \tau_3$. In Example 7.3, we found the round out and wide angle methods differed significantly at the .01 experimentwise error rate. An estimate of $\tau_1 - \tau_3$ provides us with an idea of the time saved by running wide angle as opposed to round out.

From Table 7.11 and (7.37), we obtain $Z_{12} = .05$, $Z_{13} = .125$, and $Z_{23} = .10$. From (7.38), we have

$$Z_{1.} = \frac{Z_{11} + Z_{12} + Z_{13}}{3} = \frac{0 + .05 + .125}{3} = .058,$$

$$Z_{2.} = \frac{Z_{21} + Z_{22} + Z_{23}}{3} = \frac{-.05 + 0 + .10}{3} = .017,$$

$$Z_{3.} = \frac{Z_{31} + Z_{32} + Z_{33}}{3} = \frac{-.125 - .10 + 0}{3} = -.075.$$

Note that for calculating $Z_{2.}$ and $Z_{3.}$, we have used the fact that $Z_{uv} = -Z_{vu}$.

The adjusted estimator of $\theta = \tau_1 - \tau_3$ is now obtained using (7.32) with

$$a_1 = 1, \quad a_2 = 0, \quad a_3 = -1,$$

so that from (7.40), we have

$$\tilde{\theta} = Z_{1.} - Z_{3.} = .058 - (-.075) = .133.$$

Parenthetically, it should be noted that the equivalent form (7.34) is obtained with the identifications

$$d_{11} = d_{12} = d_{13} = \frac{1}{3},$$

$$d_{21} = d_{22} = d_{23} = 0,$$

$$d_{31} = d_{32} = d_{33} = -\frac{1}{3}.$$

Table 7.11 Values of D_{uv} Differences for Data of Table 7.1

Player i	D_{12}^i	D_{13}^i	D_{23}^i
1	−.10	−.15	−.05
2	.15	.10	−.05
3	−.40	−.30	.10
4	.05	.15	.10
5	.05	.20	.15
6	−.10	−.15	−.05
7	.00	.05	.05
8	−.05	.10	.15
9	.10	.25	.15
10	.05	.15	.10
11	.05	.15	.10
12	.10	.20	.10
13	.25	.15	−.10
14	.05	.10	.05
15	.00	.10	.10
16	−.10	−.05	.05
17	.00	.20	.20
18	−.05	−.10	−.05
19	.05	.25	.20
20	.05	.25	.20
21	.05	.15	.10
22	.00	.05	.05

Comments

41. *Unadjusted Estimator*. The unadjusted estimator Z_{uv} (7.37) of Δ_{uv} (7.33) is simply the estimator associated with the sign test and previously discussed in Section 3.5. However, the Doksum adjusted estimator $\tilde{\theta}$ (7.40) is quite often different from this simple unadjusted estimator $\widetilde{Z_{uv}}$. This is the case in Example 7.5, for instance, where $Z_{13} = .125$, but $\tilde{\theta} = \widetilde{\tau_1 - \tau_3} = .133$.
42. *Ambiguities with the Unadjusted Estimators*. The unadjusted estimators Z_{uv} (7.37) lead to ambiguities in contrast estimation because they do not satisfy the linear relations that are satisfied by the contrasts they estimate. For example, $\Delta_{13} = \tau_1 - \tau_3 = (\tau_1 - \tau_2) + (\tau_2 - \tau_3) = \Delta_{12} + \Delta_{23}$, but, in general, $Z_{13} \neq Z_{12} + Z_{23}$. Thus, the two "reasonable" estimators Z_{13} and $Z_{12} + Z_{23}$ of $\Delta_{13} = \tau_1 - \tau_3$ can give different estimates. We refer to this property as the incompatibility of the unadjusted estimators Z_{uv}.
43. *Efficiency*. The adjusted estimators $\tilde{\Delta}_{uv}$ (7.39) of Δ_{uv} (7.33) are always at least as efficient as the unadjusted ones and they are compatible. They do, however, have the disadvantage that the estimator of $\Delta_{uv} = \tau_u - \tau_v$ depends on the observations from the other $k - 2$ treatments.
44. *Contrast Estimator Associated with Signed Ranks*. As noted in Comment 41, the contrast estimator $\tilde{\theta}$ (7.40) is related to paired replicates estimators associated with the sign statistic, as discussed in Sections 3.4 and 3.5. A competitor contrast estimator related to the Hodges–Lehmann paired replicates estimator associated with the signed rank statistic and utilizing between-block information is discussed in Section 7.15.

Properties

1. *Standard Deviation of $\tilde{\theta}$ (7.40)*. For the asymptotic standard deviation of $\tilde{\theta}$ (7.40), see Doksum (1967).
2. *Asymptotic Normality*. See Doksum (1967).
3. *Efficiency*. See Doksum (1967) and Section 7.16.

Problems

46. Estimate $2\tau_N - \tau_A - \tau_R$ for the metronome data of Table 7.6.

47. Illustrate, using a numerical example, the incompatibility of the unadjusted estimators Z_{uv} (see Comment 42).

48. Estimate the simple contrasts $\theta_1 = \tau_2 - \tau_1, \theta_2 = \tau_3 - \tau_1$, and $\theta_3 = \tau_3 - \tau_2$ for the CPK activity data in Table 7.3.

49. Estimate the contrast $3\tau_{ALC} - \tau_{AL} - \tau_{AC} - \tau_{LC}$ for the percentage consonants correctly identified data in Table 7.4.

50. Using the data of Table 7.4, estimate the simple contrast that represents the benefit from adding lip reading to audition in teaching severely hearing-impaired children.

51. Estimate the contrast $\tau_{AC} + \tau_{LC} - 2\tau_C$ for the percentage consonants correctly identified data in Table 7.4.

52. Estimate all contrasts found to be of interest in Problem 45 for the maximum soil temperature data in Table 7.8.

53. Estimate all possible simple contrasts for the ozone exposure data in Table 7.2.

54. Consider the percent average wind speed reduction data in Table 7.7. Use an appropriate all-treatments multiple comparison procedure (see Section 7.3) to decide which distances from the shelterbelt have significantly different reductions in average wind speed. Estimate all contrasts suggested to be important from this multiple comparison analysis.

55. Estimate the contrast $\tau_{\text{rats}} - \tau_{\text{cats}}$ for the Livesey EPT error score data of Table 7.9.

INCOMPLETE BLOCK DATA—TWO-WAY LAYOUT WITH ZERO OR ONE OBSERVATION PER TREATMENT–BLOCK COMBINATION

In two-way layout settings the most common form of data collection corresponds to the case of a single observation for every treatment–block combination. However, it is not uncommon to deal with two-way layout problems where certain treatment–block cells yield single observations, but where there are also treatment–block combinations for which we have no observations. This could be the result of a deliberate design to deal with data collection problems where it is not feasible (economically, time constraints, etc.) to collect data from every treatment–block combination or it could be simply a result of missing data from what was intended to be a complete block design.

In the next three sections we discuss procedures developed for such incomplete block data sets. In Sections 7.6 and 7.7 we present a distribution-free hypothesis test for general alternatives and an all-treatments multiple comparison procedure, respectively, for the most commonly used design specifically structured to yield less than complete block data, namely, the BIBD. In Section 7.8 we detail a distribution-free hypothesis

test for general alternatives that is applicable for two-way layout data representing an arbitrary configuration of either zero or one observation per cell.

Throughout these three sections, we continue to operate under the general conditions of Assumptions A1–A3. However, in these sections we impose the additional constraint that each c_{ij} is either 0 or 1 and $N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij} \neq kn$; that is, we have incomplete block data. We again drop the third subscript on the X variables in Sections 7.6–7.8. This will not be problematic, however, as there are no cells with more than one observation.

7.6 A DISTRIBUTION-FREE TEST FOR GENERAL ALTERNATIVES IN A RANDOMIZED BALANCED INCOMPLETE BLOCK DESIGN (BIBD) (DURBIN–SKILLINGS–MACK)

In this section we present a procedure for testing H_0 (7.2) against the general alternatives H_1 (7.3) for incomplete block data that arise from a very structured randomized BIBD. Such a BIBD corresponds to a setting where we observe $s(< k)$ treatments in each of the n blocks, every pair of treatments occurs together in the same number, λ, of blocks, and each of the k treatments is observed for a total of p times. These parameters of a BIBD must satisfy the restriction that $p(s-1) = \lambda(k-1)$, which, of course, forces additional constraints on the c_{ij}'s (see Problems 57 and 60).

Procedure

To compute the Durbin–Skillings–Mack statistic for such a balanced incomplete block design setting, we first order the available s observations from least to greatest separately within each of the n blocks. Let r_{ij} be this within-block rank of X_{ij} if there is an observation for the ith block–jth treatment combination; otherwise, let $r_{ij} = 0$. Set

$$R_j = \sum_{i=1}^{n} r_{ij}, \quad \text{for } j = 1, \ldots, k. \tag{7.42}$$

Thus, for example, R_3 is the sum (over the n blocks) of the within-blocks ranks received by the p available treatment 3 observations. (Note that each R_j will be the sum of exactly p nonzero within-blocks ranks.) The Durbin–Skillings–Mack statistic is then given by

$$\begin{aligned} D &= \left[\frac{12}{\lambda k(s+1)}\right] \sum_{j=1}^{k} \left\{R_j - \frac{p(s+1)}{2}\right\}^2 \\ &= \left\{\left[\frac{12}{\lambda k(s+1)}\right] \sum_{j=1}^{k} R_j^2\right\} - \frac{3(s+1)p^2}{\lambda}, \end{aligned} \tag{7.43}$$

where $(s+1)/2 = \sum_{j=1}^{k} r_{ij}/k$ is the average within-blocks rank assigned for each of the n blocks. Since each treatment is observed p times, it follows that $p(s+1)/2$ is the expected sum of within-blocks ranks for each of the k treatments when H_0 (7.2) is true.

To test

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } D \geq d_{\alpha,s}; \quad \text{otherwise do not reject,} \tag{7.44}$$

where the constant $d_{\alpha,s}$ is chosen to make the type I error probability equal to α. The constant $d_{\alpha,s}$ is the upper α percentile for the null $(\tau_1 = \cdots = \tau_k)$ distribution of D. Comment 49 explains how to obtain the critical values $d_{\alpha,s}$ for k treatments, n blocks, and available values of α.

Large-Sample Approximation

When H_0 is true, the statistic D has, as n tends to infinity, an asymptotic chi-square (χ^2) distribution with $k-1$ degrees of freedom. (See Comment 50 for indications of the proof.) The chi-square approximation for procedure (7.44) is

$$\text{Reject } H_0 \text{ if } D \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.45}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile of a chi-square distribution with $k-1$ degrees of freedom. To find $\chi^2_{k-1,\alpha}$, we use the `R` command `qchisq`$(1-\alpha, k-1)$. For example, to find $\chi^2_{6,.025}$, we apply `qchisq(.975,6)` and obtain $\chi^2_{6,.025} = 14.45$.

Skillings and Mack (1981) noted that this approximate procedure (7.45) can be quite conservative when α is small (say, $\leq .01$) and either the number of blocks n or the number of common occurrences λ is small. In particular, they suggest that the approximation is conservative whenever λ is not at least 3. In such cases, they strongly recommend the use of the exact values of $d_{\alpha,s}$ whenever possible.

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual treatment sums of ranks $R_1, \ldots, R_k$. In such cases, the significance level associated with procedure (7.44) is only approximately equal to α. (See Comment 51 for discussion of how to construct a conditionally distribution-free test of H_0 even when there are ties within some of the blocks.)

EXAMPLE 7.6 ***Chemical Toxicity.***

Moore and Bliss (1942) compared the toxicity of each of seven chemicals applied to *Aphis rumicis*, a black aphid found on nasturtiums. The logarithm of the dose required to kill 95% of the insects exposed to a chemical was the measurement reported. Since the experimenters could test only three chemicals in any given day, they used a balanced incomplete block design requiring 7 days for completion of the experiment. The toxicities for the studied chemicals are shown in Table 7.12.

Table 7.12 Logarithm of Toxic Dosages

	Chemical						
Day	A	B	C	D	E	F	G
1	.465	.343		.396			
2	.602		.873		.634		
3			.875	.325			.330
4	.423					.987	.426
5		.652	1.142			.989	
6		.536			.409		.309
7				.609	.417	.931	

Source: W. Moore and C. I. Bliss (1942).

This experiment constitutes a BIBD with $k = 7$ treatments, of which $s = 3$ are observed in each of the $n = 7$ blocks, every pair of treatments occur together in $\lambda = 1$ of the blocks, and each of the treatments is observed for a total of $p = 3$ times. We are interested in assessing whether there are any differences in the toxicities of the seven chemicals relative to *A. rumicis*. We will use the approximate procedure (7.45) to test if there are any differences in the toxicities of the seven chemicals. For the sake of illustration, we take the approximate significance level to be $\alpha = .05$. Using the R command `qchisq`$(1-\alpha,\ k)$, we find the value $\chi^2_{6,.05} =$ `qchisq(.95, 6)` $= 12.59$ and procedure (7.45) reduces to

$$\text{Reject } H_0 \text{ if } D \geq 12.59.$$

Now, we illustrate the computations leading to the sample value of D (7.43). Ranking from 1 to 3 within each of the seven blocks (days) and summing across the blocks for each of the chemicals, we obtain the following treatment sums of ranks:

$$R_1 = 3+1+1 = 5, \quad R_2 = 1+1+3 = 5, \quad R_3 = 3+3+3 = 9, \quad R_4 = 2+1+2 = 5,$$

$$R_5 = 2+2+1 = 5, \quad R_6 = 3+2+3 = 8, \quad R_7 = 2+2+1 = 5.$$

Hence, from (7.43), we find that

$$D = \left\{\left[\frac{12}{1(7)(4)}\right](5^2+5^2+9^2+5^2+5^2+8^2+5^2)\right\} - \frac{3(4)(3^2)}{1}$$

$$= \left\{\frac{3(270)}{7}\right\} - 108 = 7.71.$$

Comparison of this observed value of 7.71 with the approximate critical value $\chi^2_{6,.05} = 12.59$ leads us to conclude that there is not strong sample evidence to indicate any significant difference between the seven studied chemicals with respect to their toxicities for *A. rumicis*. In fact, the observed value of $D = 7.71$ is approximately the .26 upper percentile for the chi-square distribution with 6 degrees of freedom (i.e., `qchisq(.74, 6)` ≈ 7.71). Thus, the approximate P-value for these data and procedure

(7.45) is .26, providing further evidence of the similarity of the studied chemicals with respect to their toxic effects on *A. rumicis*.

Comments

45. *More General Setting*. We could replace Assumptions A1–A3 and H_0 (7.2) with the more general null hypothesis that all possible $(s!)^n$ rank configurations for the nonzero r_{ij}'s are equally likely. Procedure (7.44) remains distribution-free for this more general hypothesis.
46. *Design Rationale*. In a two-way layout setting with no replications within block–treatment combinations, it is best to use a randomized complete block design (as discussed in Sections 7.1–7.5) whenever possible. However, there are times when experimental constraints such as fixed costs or limited time or facilities make it impossible to obtain an observation for every treatment–block combination. When this is the case, the use of a balanced incomplete block design is often a good alternative. Such a BIBD provides for sufficient data to be collected to permit comparison of each treatment with every other one. Moreover, the fact that the BIBD imposes a rigid structure on the missing observations within and across the blocks enables the associated data analysis to be both relatively simple and efficient.
47. *Motivation for the Test*. Under Assumptions A1–A3 and H_0 (7.2), each of the block rank vectors $\boldsymbol{R}_i^*$ for those s observations present in the ith block, $i = 1, \ldots, n$, has a uniform distribution over the set of all $s!$ permutations of the vector of integers $(1, 2, \ldots, s)$. If r_{ij} is nonzero, it follows that $E_0(r_{ij}) = (1/s!)[(s-1)!] \sum_{t=1}^{k} t = (s+1)/2$, the average rank being assigned separately to the partial data in each of the blocks. Thus, $E_0(R_j) = \sum_{i=1}^{n} E_0(r_{ij}) = p(s+1)/2$, for $j = 1, \ldots, k$, because there are observations in exactly p of the blocks for each of the k treatments. Therefore, we would expect each of the R_j's to be close to $p(s+1)/2$ when H_0 is true. Since the test statistic D (7.43) is a constant times a sum of squared differences between the observed treatment sums of ranks, R_j, and their common null expected value, $E_0(R_j) = p(s+1)/2$, small values of D represent agreement with H_0 (7.2). When the τ's are not all equal, we would expect a portion of the associated treatment sums of ranks to differ from their common null expectation, $p(s+1)/2$, with some tending to be smaller and some larger. The net result (after squaring the observed differences to obtain the $[R_j - p(s+1)/2]^2$ terms) would be a large value of D. This quite naturally suggests rejecting H_0 in favor of H_1 (7.3) for large values of D and motivates procedures (7.44) and (7.45).
48. *Assumptions.* We emphasize that Assumption A3 stipulates that the ns cell distributions F_{ij} for those treatment–block combinations where observations are collected can differ at most in their locations (medians) and that these location differences (if any) must be a result of additive block and/or treatment effects (i.e., there is no interaction between the treatment and block factors). In particular, Assumption A3 requires that the ns underlying distributions belong to the same general family (F) and that they do not differ in scale parameters (variability). We do note, however, that the test procedure (7.44) remains distribution-free under the less restrictive setting where Assumption A3 is replaced by the weaker condition

A3′. The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F_i(u - \tau_j), -\infty < u < \infty,$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where $F_1, \ldots, F_n$ are arbitrary distribution functions for continuous distributions with unknown medians $\theta_1, \ldots, 0_n$, respectively, and, as before, τ_j is the unknown additive treatment effect contributed by the jth treatment.

Assumption A3 then corresponds to Assumption A3′ with the additional condition that $F_1 \equiv \cdots \equiv F_n$ (see also Comment 45).

49. *Derivation of the Distribution of D under H_0 (No-Ties Case).* The null distribution of D (7.43) can be obtained using the fact that under H_0 (7.2), all possible $(s!)^n$ rank configurations for the nonzero r_{ij}'s are equally likely. We now take the simplest (but not very useful in practice) balanced incomplete block design corresponding to $k = 3, s = 2, n = 3, p = 2$, and $\lambda = 1$ to illustrate how the null distribution can be derived. In this case, D (7.43) reduces to $D = (\frac{4}{3})R^* - 36$, where $R^* = R_1^2 + R_2^2 + R_3^2$.

The value of D for each of the $(2!)^3 = 8$ possible rank configurations for this setting are presented below.

I	II	III
1	2	
1		2
	1	2

$R^* = 29, D = 2.67$

I	II	III
2	1	
1		2
	1	2

$R^* = 29, D = 2.67$

I	II	III
1	2	
2		1
	1	2

$R^* = 27, D = 0$

I	II	III
2	1	
2		1
	1	2

$R^* = 29, D = 2.67$

I	II	III
1	2	
1		2
	2	1

$R^* = 29, D = 2.67$

I	II	III
1	2	
2		1
	2	1

$R^* = 29, D = 2.67$

I	II	III
2	1	
1		2
	2	1

$R^* = 27, D = 0$

I	II	III
2	1	
2		1
	2	1

$R^* = 29, D = 2.67$

Thus, we find

$$P_0\{D = 0\} = .25 \quad \text{and} \quad P_0\{D = 2.67\} = .75.$$

Note that we have derived the null distribution of D without specifying the common form (F) of the underlying distribution function for the X's under H_0 beyond the point of requiring that it be continuous. This is why the test procedure (7.44) based on D is called a *distribution-free Procedure*. From the

null distribution of D, we can determine the critical value $d_{\alpha,s}$ and control the probability α of falsely rejecting H_0 when it is true, and this error probability does not depend on the specific form of the common underlying continuous X distribution.

For a given BIBD design with incidence matrix **obs.mat**, the R command cDurSkiMa(α, **obs.mat**) can be used to find the available upper-tail critical values $d_{\alpha,s}$ for possible values of D. The incidence matrix will be an $n \times k$ matrix of ones and zeroes, which indicate where the data are observed and unobserved, respectively. Methods for finding the incidence matrix for various BIBD designs are given in the literature. While the incidence matrix will not be unique for a given (k, n, s, λ, p) combination, the distribution of D under H_0 will be the same for any of the possible incidence matrices. For a given available significance level α, the critical value $d_{\alpha,s}$ then corresponds to $P_0(D \geq d_{\alpha,s}) = \alpha$ and is given by cDurSkiMa(α, **obs.mat**) $= d_{\alpha,s}$. Thus, for example, for the BIBD combination $(k, n, s, \lambda, p) = (3, 3, 2, 1, 2)$, one possibility would be $\textbf{obs.mat} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, which would yield cDurSkiMa(.75, **obs.mat**) $= d_{.75,s} = 2.67$, as noted previously in this comment. As a second more practical example, we can consider the case of $(k, n, s, \lambda, p) = (6, 15, 4, 6, 10)$. A possible incidence matrix is given by

$$\textbf{obs.mat} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix},$$

which yields cDurSkiMa(.0487, **obs.mat**) $= d_{.0487,s} = 10.80$.

50. *Large-Sample Approximation*. Define the random variables $T_j = R_j - E_0(R_j) = R_j - p(s+1)/2$, for $j = 1, \ldots, k$. Since each R_j is a sum, it is not surprising (see, e.g., Skillings and Mack (1981) for formal justification) that a properly standardized version of the vector $\boldsymbol{T}^* = (T_1, \ldots, T_{k-1})$ has an asymptotic (n tending to infinity) $(k-1)$-variate normal distribution with mean vector $\mathbf{0} = (0, \ldots, 0)$ and appropriate covariance matrix Σ when the null hypothesis H_0 is

true. (Note that $\boldsymbol{T}^*$ does not include $T_k = R_k - p(s+1)/2$, because T_k can be expressed as a linear combination of $T_1, \ldots, T_{k-1}$. This is the reason that the asymptotic normal distribution is $(k-1)$-variate and not k-variate.) Since the test statistic D (7.43) is a quadratic form in the variables $(T_1, \ldots, T_{k-1})$, it is therefore quite natural that D has an asymptotic (n tending to infinity) chi-square distribution with $k-1$ degrees for freedom.

51. *Exact Conditional Distribution of D with Ties among the Observed X Values within Blocks*. To have a test with exact significance level even in the presence of tied X's within some of the blocks, we need to consider all $(s!)^n$ block rank configurations, where now these within-blocks ranks are obtained using average ranks to break ties. As in Comment 49, it still follows that under H_0 each of the $(s!)^n$ block rank configurations (now with these tied ranks) is equally likely. For each such configuration, the value of D is computed and the results are tabulated. We illustrate this for the same setting as was used in the untied example in Comment 49 (namely, $k=3$, $s=2$, $n=3$, $p=2$, and $\lambda=1$), except here we assume that the two observations within block one are tied in value. Thus, the block ranks we are dealing with here are (1.5, 1.5), (1, 2), and (1, 2) for blocks 1, 2, and 3, respectively.

As in Comment 49, D (7.43) reduces to $D = (\frac{4}{3})R^* - 36$, where $R^* = R_1^2 + R_2^2 + R_3^2$. Since the rank configuration for the first block is always (1.5, 1.5), we need only compute the value of D for $(2!)^2 = 4$ possible rank configurations. The values of D for these four configurations are as follows:

I	II	III
1.5	1.5	
1		2
	1	2

$R^* = 28.5, D = 2$

I	II	III
1.5	1.5	
1		2
	2	1

$R^* = 27.5, D = 0.67$

I	II	III
1.5	1.5	
2		1
	1	2

$R^* = 27.5, D = 0.67$

I	II	III
1.5	1.5	
2		1
	2	1

$R^* = 28.5, D = 2$

Thus, we find

$$P_0\{D = 0.67\} = .50 \quad \text{and} \quad P_0\{D = 2\} = .50.$$

This distribution is called the *conditional distribution* or the *permutation distribution* of D, given the set of tied within-blocks ranks (1.5, 1.5), (1, 2), and (1, 2).

52. *Historical Development*. The test procedures (7.44) and (7.45) based on D were first proposed by Durbin (1951). Later, Skillings and Mack (1981) studied a more general procedure for arbitrary incomplete block data (see Section 7.8)

and, for the first time, made available the exact critical values $d_{\alpha,s}$ for procedure (7.44) for a reasonable set of balanced incomplete block designs.

Properties

1. *Consistency*. See van Elteren and Noether (1959).
2. *Asymptotic Chi-Squaredness*. See Durbin (1951), Benard and van Elteren (1953) or Skillings and Mack (1981).
3. *Efficiency*. See van Elteren and Noether (1959) and Section 7.16.

Problems

56. Mendenhall (1968) discusses an experiment that was conducted to compare the effects of seven different chemical substances on the skin of male rats. The necessity to use relatively homogeneous patches of a rat's skin for the study restricted the experimenter to three experimental units (patches of skin) per animal. However, to avoid the confounding effect of rat-to-rat variability in the comparison of the seven chemicals, the experimenter was obligated to block on rats and any given rat could be treated with only three of the seven chemicals. This resulted in the use of a balanced incomplete block design with parameters $k = 7$, $n = 7$, $s = 3$, $p = 3$, and $\lambda = 1$. The experimental measurements for the study are presented in Table 7.13.

Apply procedure (7.45) with approximate significance level $\alpha \approx .05$ to these data to test the hypothesis of interest.

57. Verify that the relationship $p(s-1) = \lambda(k-1)$ must hold for a balanced incomplete block design.

58. Verify that the two representations for D (7.43) are, in fact, equivalent.

59. What are the maximum and minimum values for the test statistic D (7.43)? What rank configurations lead to these maximum and minimum values?

60. Consider the relationship $p(s-1) = \lambda(k-1)$ that must hold for a balanced incomplete block design. What constraints does this condition place on the c_{ij}'s for the data?

61. Kuehl (1994) described an experiment by J. Berry and A. Deutschman at the University of Arizona designed to study the effect of pressure on percent conversion of methyl glucoside to monovinyl isomers. The conversion is achieved by addition of acetylene to methyl glucoside in the presence of a base under high pressure. Five pressures were of interest in the study, but only three could be examined at any one time under identical experimental conditions because

Table 7.13 Reactions of Male Rats to Chemical Substances

	Chemical Substance						
Rat	A	B	C	D	E	F	G
1	10.2	6.9		14.2			
2			9.9	12.9		14.1	
3		12.1	11.7		8.6		
4				14.3	9.1		7.7
5		8.8				16.3	8.6
6	13.1				9.2	15.2	
7	11.3		9.7				6.2

Source: W. Mendenhall (1968).

Table 7.14 Percent Conversion of Methyl Glucoside to Monovinyl Isomers

	Pressure (psi)				
Experimental run	250	325	400	475	550
1	16	18		32	
2	19			46	45
3		26	39		61
4			21	35	55
5		19		47	48
6	20		33	31	
7	13	13	34		
8	21		30		52
9	24	10			50
10		24	31	37	

Source: R. O. Kuehl (1994).

of limited laboratory space. This necessitated the use of a balanced incomplete block design. The data obtained in the experiment and design are given in Table 7.14.

State the parameters for the BIBD employed in this chemical conversion study. Apply procedure (7.44) to these data to assess whether pressure (at the levels included in the study) has any effect on the percent conversion of methyl glucoside to monovinyl isomers.

62. Consider the BIBD corresponding to $k = 5$, $n = 10$, $s = 3$, $p = 6$, and $\lambda = 3$. Compare the critical region for the exact level $\alpha = .0499$ test of H_0 (7.2) based on D with the critical region for the corresponding nominal level $\alpha = .0499$ test based on the large-sample approximation.

63. Consider the BIBD corresponding to $k = 4$, $n = 6$, $s = 2$, $p = 3$, and $\lambda = 1$. Obtain the form of the exact null H_0 distribution of D (7.43) for the case of no-tied observations.

64. Consider the BIBD corresponding to $k = 5$, $n = 20$, $s = 2$, $p = 8$, and $\lambda = 2$. Compare the critical region for the exact level $\alpha = .0685$ test of H_0 (7.2) based on D with the critical region for the corresponding nominal level $\alpha = .0685$ test based on the large-sample approximation.

65. Consider the BIBD corresponding to $k = 4$, $n = 6$, $s = 2$, $p = 3$, and $\lambda = 1$. Suppose that the two observations in each of blocks 3 and 5 are tied. Obtain the conditional exact probability distribution of D under H_0 (7.2) when average ranks are used to break these two sets of within-blocks ties. Compare this conditional null distribution of D with the null distribution for D obtained in Problem 63 when there are no ties.

66. Consider the percentage consonants correctly identified data in Table 7.4 for conditions AL, AC, LC, and ALC only. Suggest a possible BIBD that could have been utilized in this study for these four treatments to reduce the number of conditions under which each severely hearing-impaired child had to be observed. Using a random mechanism for deciding how to apply the BIBD in question to the existing data set, analyze the corresponding data subset to assess whether there are any differences in the effectiveness of the conditions AL, AC, LC, and ALC for teaching severely hearing-impaired children.

67. Consider the percentage consonants correctly identified data in Table 7.4 for conditions A, L, and C only, Suggest a possible BIBD that could have been utilized in this study for these three treatments to reduce the number of conditions under which each severely hearing-impaired child had to be observed. Using a random mechanism for deciding how to apply the BIBD in question to the existing data set, analyze the corresponding data subset to assess whether there are any differences in the effectiveness of the conditions A, L, and C for teaching severely hearing-impaired children.

7.7 ASYMPTOTICALLY DISTRIBUTION-FREE TWO-SIDED ALL-TREATMENTS MULTIPLE COMPARISONS FOR BALANCED INCOMPLETE BLOCK DESIGNS (SKILLINGS–MACK)

In this section we present an asymptotically distribution-free multiple comparison procedure using the Friedman within-blocks ranks that is designed to make decisions about individual differences between pairs of treatment effects (τ_i, τ_j), for $i < j$, for data obtained from a balanced incomplete block design. The multiple comparison procedure of this section would generally be applied to BIBD data *after* rejection of H_0 (7.2) with the Durbin–Skillings–Mack procedure from Section 7.6. In this setting we will reach conclusions about all $k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two-sided in nature.

Procedure

Let $R_1, \ldots, R_k$ be the treatment sums of within-blocks ranks given by (7.42). Calculate the $k(k-1)/2$ absolute differences $|R_u - R_v|$, $1 \le u < v \le k$.

When H_0 (7.2) is true, the $k(k-1)/2$-component vector $(R_1, \ldots, R_k)$ has, when properly standardized and as n tends to infinity, an asymptotic $(k-1)$-variate normal distribution with appropriate mean vector and covariance matrix (see Skillings and Mack (1981) for details of the proof). At an approximate experimentwise error rate of α, the Skillings–Mack two-sided all-treatments multiple comparison procedure reaches its $k(k-1)/2$ pairwise decisions, corresponding to each (τ_u, τ_v) pair, $1 \le u < v \le k$, by the criterion

$$\begin{aligned} &\text{Decide } \tau_u \neq \tau_v \text{ if } |R_u - R_v| \ge [(s+1)(ps - p + \lambda)/12]^{1/2} q_\alpha; \\ &\qquad \text{otherwise decide } \tau_u = \tau_v, \end{aligned} \tag{7.46}$$

where q_α is the upper αth percentile for the distribution of the range of k independent $N(0, 1)$ variables. To find q_α for k treatments and a specified experimentwise error rate α, we use the R command `cRangeNor(`α`,`k`)`. For example, to find $q_{.025}$ for $k = 6$ treatments, we apply `cRangeNor(.025,6)` and obtain $q_{.025} = 4.361$ for $k = 6$. (See also Comment 55.)

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual treatment sums of ranks $R_1, \ldots, R_k$.

EXAMPLE 7.7 ***Chemical Toxicity.***

For the sake of illustration, we apply procedure (7.46) to the chemical toxicity data relative to the black aphid, *A. rumicis*, as previously discussed in Example 7.6, even though the Durbin–Skillings–Mack procedure did not find any significant differences (approximate P-value of .26) between the treatment effects. Taking our approximate

experimentwise error rate to be $\alpha \approx .05$, we find `cRangeNor(.05, 7)` $= q_{.05} = 4.170$ for $k = 7$ and procedure (7.46) reduces to

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |R_u - R_v| \geq [4(9-3+1)/12]^{1/2}(4.170) = 6.370.$$

Using the treatments sums of within-blocks ranks obtained in Example 7.6, we find that

$$\begin{array}{lllll} |R_2 - R_1| = 0, & |R_3 - R_1| = 4, & |R_4 - R_1| = 0, & |R_5 - R_1| = 0, & |R_6 - R_1| = 3, \\ |R_7 - R_1| = 0, & |R_3 - R_2| = 4, & |R_4 - R_2| = 0, & |R_5 - R_2| = 0, & |R_6 - R_2| = 3, \\ |R_7 - R_2| = 0, & |R_4 - R_3| = 4, & |R_5 - R_3| = 4, & |R_6 - R_3| = 1, & |R_7 - R_3| = 4, \\ |R_5 - R_4| = 0, & |R_6 - R_4| = 3, & |R_7 - R_4| = 0, & |R_6 - R_5| = 3, & |R_7 - R_5| = 0, \\ & & |R_7 - R_6| = 3. & & \end{array}$$

Since all these absolute differences are less than the critical value, 6.370, we see that the $7(6)/2 = 21$ decisions at this approximate experimentwise error rate of .05 are that $\tau_u = \tau_v$, for $1 \leq u < v \leq 7$. This is, of course, not at all surprising, because, in Example 7.6, the Durbin–Skillings–Mack test procedure found no support for rejecting H_0 (7.2) for these data.

Comments

53. *Rationale for Multiple Comparison Procedure.* The rationale behind the multiple comparison procedure of this section for data from a balanced incomplete block design is similar to that behind the two-sided multiple comparison procedures for data from a complete randomized block design. For further discussion, see Comment 24.
54. *Experimentwise Error Rate.* The use of an experimentwise error rate represents a very conservative approach to multiple comparisons. We are insisting that the probability of making only correct decisions be $1 - \alpha$ when the null hypothesis H_0 (7.2) of treatment equivalence is true. Thus, we have a high degree of protection when H_0 is true, but we often apply such techniques when we have evidence (perhaps based on a priori information or perhaps obtained by applying the Durbin–Skillings–Mack test, as in Example 7.7) that H_0 is not true. The protection under H_0 also makes it harder for the procedure to judge treatments as differing significantly when, in fact, H_0 is false, and this difficulty becomes more severe as k increases. See Comment 6.54 for additional discussion of experimentwise error rates.
55. *Conservative Procedure.* Skillings and Mack (1981) also proposed a conservative multiple comparison procedure that guarantees an upper bound on the experimentwise error rate. Let $R_1, \ldots, R_k$ be the treatment sums of within-blocks ranks given by (7.42). At an experimentwise error rate *no greater* than α, the Skillings–Mack conservative two-sided all-treatments multiple comparison procedure reaches its $k(k-1)/2$ decisions through the criterion

$$\begin{aligned} &\text{Decide } \tau_u \neq \tau_v \text{ if } |R_u - R_v| \geq [k\lambda d_{\alpha,s}(s+1)/6]^{1/2}; \\ &\qquad \text{otherwise decide } \tau_u = \tau_v, \end{aligned} \tag{7.47}$$

where $d_{\alpha,s}$ is the upper α percentile for the null distribution of the Durbin–Skillings–Mack statistic D (7.43). Skillings and Mack (1981) note that although procedure (7.47) does not require a large number of blocks, it is, nevertheless, rather conservative because it is based on the projection procedure of Scheffé; that is, the true experimentwise error rate might be considerably smaller than the bound α provided by (7.47). As a result, they recommend using the approximation (7.46) whenever the number of blocks is reasonably large.

56. *Contrast Estimators for BIBD's.* Greenberg (1966) proposed a method of contrast estimation for general (including balanced) incomplete block designs where the number of observations in a block is smaller than the number of treatments to be compared.

57. *Dependence on Observations from Other Noninvolved Treatments.* The all-treatments multiple comparison procedure of this section suffers from the same disadvantage as do the other two-way layout multiple comparison procedures of this chapter. The decision between treatment u and treatment v can be affected by changes only in the observations from one or more of the other $k - 2$ treatments that are not directly involved.

Properties

1. *Asymptotic Multivariate Normality*. See Skillings and Mack (1981).
2. *Efficiency.* See Section 7.16.

Problems

68. Apply procedure (7.46) to the chemical substance effect data of Table 7.13 in Problem 56.

69. Illustrate the difficulty discussed in Comment 57 by means of a numerical example.

70. Apply procedure (7.46) to the percent conversion data of Table 7.14 in Problem 61.

71. Consider the chemical toxicity data of Table 7.12 in Example 7.6. Find the smallest approximate experimentwise error rate at which the most significant difference(s) in black aphid (*A. rumicis*) toxicity between the studied substances would be detected by procedure (7.46).

72. Consider the chemical substance effect data of Table 7.13 in Problem 56. Find the smallest approximate experimentwise error rate at which procedure (7.46) would declare that chemical substances F and G have differing effects on the skin of male rats.

73. Consider the percent conversion data of Table 7.14 in Problem 61. Find the smallest approximate experimentwise error rate at which the most significant difference in the effects of the various pressures on the percent conversion of methyl glucoside to monovinyl isomers would be detected by procedure (7.46).

7.8 A DISTRIBUTION-FREE TEST FOR GENERAL ALTERNATIVES FOR DATA FROM AN ARBITRARY INCOMPLETE BLOCK DESIGN (SKILLINGS–MACK)

Not every set of data resulting from less than a randomized complete block design satisfies the necessary constraints (see Section 7.6) to be analyzed by the Durbin–Skillings–Mack procedure for balanced incomplete block designs. In this section we present a general

procedure for analyzing data from a two-way layout where there are either zero or one observation for each treatment–block combination but where there is not necessarily any nice pattern to the particular combinations for which we do not have observations. Such an incomplete data configuration could, of course, be intentionally designed this way, but it could also be the consequence of missing observations from an experiment that was intended to yield data for a randomized complete block design.

For this general two-way layout setting, let s_i denote the number of treatments for which an observation is present in block i, for $i = 1, \ldots, n$. (If $s_i = 1$ for block i, we remove that block from the analysis. Therefore, throughout this section, n will denote the number of blocks for which $s_i \geq 2$.) We discuss a distribution-free procedure for testing H_0 (7.2) against the general alternatives H_1 (7.3) when we are faced with such arbitrarily incomplete block data.

Procedure

To compute the Skillings–Mack statistic for arbitrarily incomplete block data, we first rank the s_i observed data values in block i from least to greatest, for each block $i = 1, \ldots, n$. Thus, in the ith block, we will be assigning ranks $1, 2, \ldots, s_i$. For $i = 1, \ldots, n$ and $j = 1, \ldots, k$, let

$$\begin{aligned} r_{ij} &= \text{rank of } X_{ij} \text{ among the observations in block } i, \quad \text{if } c_{ij} = 1, \\ &= (s_i + 1)/2, \quad \text{if } c_{ij} = 0, \end{aligned}$$

where $(s_i + 1)/2$ is the average of the ranks assigned to the observations present in the ith block. Set

$$A_j = \sum_{i=1}^{n} \left(\frac{12}{s_i + 1} \right)^{1/2} \left(r_{ij} - \frac{s_i + 1}{2} \right), \quad j = 1, \ldots, k. \tag{7.48}$$

Thus, A_j is the weighted sum of centered (around the block average) ranks for the observations from the jth treatment, with the block weighting factor $[12/(s_i + 1)]^{1/2}$ being inversely proportional to the square root of the number of observations present in the block (see Comment 63). Set

$$\boldsymbol{A} = (A_1, \ldots, A_{k-1}). \tag{7.49}$$

(Without loss of generality, we have chosen to omit A_k from the vector $\boldsymbol{A}$. The A_j's are linearly dependent, because a weighted linear combination of all k of them is a constant. We could omit any one of the A_j's in the definition of $\boldsymbol{A}$ (7.49), and the procedure we now describe would lead to the same value of the test statistic. For further discussion, see Skillings and Mack (1981).)

The covariance matrix for $\boldsymbol{A}$ under H_0 (7.2) is given by

$$\Sigma_0 = \begin{bmatrix} \sum_{t=2}^{k} \lambda_{1t} & -\lambda_{12} & -\lambda_{13} & \cdots & -\lambda_{1,k-1} \\ -\lambda_{12} & \sum_{t \neq 2}^{k} \lambda_{2t} & -\lambda_{23} & \cdots & -\lambda_{2,k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ -\lambda_{1,k-1} & -\lambda_{2,k-1} & -\lambda_{3,k-1} & \cdots & \sum_{t \neq k-1}^{k} \lambda_{k-1,t} \end{bmatrix} \tag{7.50}$$

where, for $t \neq q = 1, \ldots, k$,

$$\lambda_{qt} = \lambda_{tq} = [\text{number of blocks in which both treatments } q \text{ and } t \text{ are observed}]. \tag{7.51}$$

Let Σ_0^- be any (see Comment 62) generalized inverse for Σ_0. The Skillings–Mack statistic is then given by

$$SM = \boldsymbol{A}\Sigma_0^- \boldsymbol{A}'. \tag{7.52}$$

(We note that if $\lambda_{qt} > 0$ for all $q \neq t$, then the rank of the covariance matrix Σ_0 (7.50) is $k - 1$, and we can simply use the ordinary inverse Σ_0^{-1} in the definition of SM (7.52).)

To test

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \tau_2, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } SM \geq sm_\alpha; \quad \text{otherwise do not reject,} \tag{7.53}$$

where the constant sm_α is chosen to make the type I error probability equal to α. Comment 64 explains how to obtain the critical values sm_α for a two-way layout configuration with k treatments, n blocks, and observation indicators $c_{ij}, i = 1, \ldots, n; j = 1, \ldots, k$.

Large-Sample Approximation

When H_0 (7.2) is true and $\lambda_{qt} > 0$ for every $q \neq t = 1, \ldots, k$ (i.e., every pair of treatments occur together in at least one block), the statistic SM has, as n tends to infinity, an asymptotic chi-square (χ^2) distribution with $k - 1$ degrees of freedom (see Comment 65 for indications of the proof). The chi-square approximation for procedure (7.53) is

$$\text{Reject } H_0 \text{ if } SM \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.54}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile of a chi-square distribution with $k - 1$ degrees of freedom. To find $\chi^2_{k-1,\alpha}$, we use the R command `qchisq`$(1 - \alpha, k - 1)$. For example, to find $\chi^2_{5,.05}$, we apply `qchisq(.95,5)` and obtain $\chi^2_{5,.05} = 11.071$.

As with the BIBD procedure discussed in Section 7.6, Skillings and Mack (1981) have pointed out that this approximate procedure (7.54) can be quite conservative when α is small (say, $\leq .01$) and the number of blocks, n, is not large. In such cases, it is preferable to generate the exact critical value sm_α and use procedure (7.53). (See Comment 64.) (We should also point out that the approximate procedure (7.54) is simply not applicable if there are at least two treatments that do not have observations together in any of the blocks; that is, if $\lambda_{qt} = 0$ for at least one pair $q \neq t = 1, \ldots, k$.)

Table 7.15 Subset of Data on the Influence of Rhythmicity of Metronome on Speech Fluency

	Dysfluencies under each condition		
Subject	*R*	*A*	*N*
1	3(1)	5(2)	15(3)
2	1(1)	3(2)	18(3)
3	5(2)	4(1)	21(3)
4	2(1)	—	6(2)
5	0(1)	2(2)	17(3)
6	0(1)	2(2)	10(3)
7	0(1)	3(2)	8(3)
8	0(1)	2(2)	13(3)

Source: J. P. Brady (1969).

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual treatment weighted sums of centered ranks $A_1, \ldots, A_k$. In such cases, the significance level associated with procedure (7.53) is only approximately equal to α. (See Comment 66 for discussion of how to construct a conditionally distribution-free test of H_0 even when there are tied observations within some of the blocks.)

EXAMPLE 7.8 ***Effect of Rhythmicity of a Metronome on Speech Fluency.***

Consider Table 7.15, which contains a subset (with subject labels renumbered) of the data in Table 7.6 obtained by Brady (1969) in his study of the influence of rhythmicity of a metronome on the speech of stutterers, where the missing observation for subject 4 might be due to a malfunction in the arrhythmic metronome during her evaluation.

Here we have a two-way layout with $k = 3$, $n = 8$, and all $c_{ij's} = 1$ except $c_{42} = 0$. We illustrate the computations leading to the sample value of *SM*. The within-blocks (subjects) ranks (r_{ij}'s) for the observations present are also given in Table 7.15 in parentheses after the data values. With respect to the missing value for subject 4 under condition *A*, the average rank $(2 + 1)/2 = 1.5$ for subject 4 is assigned as the value for r_{42}. From (7.48), we compute the weighted sums of centered ranks to be

$$A_1 = \left\{ \left[\frac{12}{(3+1)} \right]^{1/2} [(1-2) + (1-2) + (2-2) + (1-2) + (1-2) + (1-2) + (1-2)] + \left[\frac{12}{(2+1)} \right]^{1/2} (1 - 1.5) \right\}$$

$$= 1.732(-6) + 2(-.5) = -11.392,$$

$$A_2 = \left\{ \left[\frac{12}{(3+1)} \right]^{1/2} [(2-2) + (2-2) + (1-2) + (2-2) + (2-2)$$

$$+ (2-2) + (2-2)] + \left[\frac{12}{(2+1)}\right]^{1/2} (1.5 - 1.5)\Bigg\}$$

$$= 1.732(-1) + 2(0) = -1.732,$$

and

$$A_3 = \Bigg\{\left[\frac{12}{(3+1)}\right]^{1/2} [(3-2) + (3-2) + (3-2) + (3-2) + (3-2)$$

$$+ (3-2) + (3-2)] + \left[\frac{12}{(2+1)}\right]^{1/2} (2 - 1.5)\Bigg\}$$

$$= 1.732(7) + 2(.5) = 13.124.$$

Thus, we obtain $\boldsymbol{A} = (-11.392, -1.732)$.

With the single missing observation for subject 4 under condition A, the combination counts λ_{qt} (7.51) are $\lambda_{12} = 7$, $\lambda_{13} = 8$, and $\lambda_{23} = 7$. Hence, from representation (7.50), the null covariance matrix Σ_0 has form

$$\Sigma_0 = \begin{bmatrix} 15 & -7 \\ -7 & 14 \end{bmatrix}.$$

Since each of λ_{12}, λ_{13}, and λ_{23} is positive, the rank of Σ_0 is 2, and its ordinary inverse is

$$\Sigma_0^{-1} = \left(\tfrac{1}{161}\right) \begin{bmatrix} 14 & 7 \\ 7 & 15 \end{bmatrix} = \begin{bmatrix} .0870 & .0435 \\ .0435 & .0932 \end{bmatrix}.$$

From (7.52), we obtain

$$\begin{aligned} SM &= \boldsymbol{A}\Sigma_0^{-1}\boldsymbol{A}' \\ &= (-11.392, -1.732) \begin{bmatrix} .0870 & .0435 \\ .0435 & .0932 \end{bmatrix} \begin{pmatrix} -11.392 \\ -1.732 \end{pmatrix} \\ &= 13.287. \end{aligned}$$

For a given arbitrary incomplete block design with incidence matrix **obs.mat**, the R command `cSkilMack`(α,**obs.mat**) can be used to find the available upper-tail critical values sm_α for possible values of SM. The incidence matrix is an $n \times k$ matrix of ones and zeroes, which indicate where the data are observed and unobserved, respectively. For a given available significance level α, the critical value sm_α then corresponds to $P_0(SM \geq sm_\alpha) = \alpha$ and is given by `cSkilMack`(α, **obs.mat**) $= sm_\alpha$. For this example,

we define the incidence matrix as

$$\texttt{obs.mat} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

and find `cSkilMack(.0097,obs.mat)` $= sm_{.0097} = 8.528$. Since we have observed $SM = 13.287 > sm_{.0097} = 8.528$. the P-value for this test procedure is smaller than .0097. (In fact, using `pSkilMack(x)` where **x** is the metronome data, we find that the P-value for this test is .00006.) Thus, there is strong evidence to support the hypothesis that the rhythmicity of a metronome does, indeed, influence the speech of stutterers (see also Comment 60 and Problem 76).

Comments

58. *More General Setting.* We could replace Assumptions A1–A3 and H_0 (7.2) with the more general null hypothesis that all possible $\prod_{i=1}^{n} s_i!$ rank configurations for the within-blocks ranks, r_{ij}, of the *observed* data are equally likely. Procedure (7.53) remains distribution-free for this more general hypothesis.

59. *Motivation for the Test.* Under Assumptions A1–A3 and H_0 (7.2), the block rank vector R_i^* for those s_i observations present in the ith block has a uniform distribution over the set of all $s_i!$ permutations of the vector of integers $(1, 2, \ldots, s_i)$, and this is true for all blocks, $i = 1, \ldots, n$. For those r_{ij} corresponding to observations present in the collected data set, it is then the case that $E_0(r_{ij}) = (s_i + 1)/2$, the average rank being assigned to the partial data present in the ith block, for every block $i = 1, \ldots, n$. Thus, it follows from (7.48) and the definition of r_{ij} for an empty cell that $E_0(A_j) = 0$ for every $j = 1, \ldots, k$. Therefore, we would expect each of the A_j's to be close to zero when H_0 is true. Since the test statistic SM (7.52) is a quadratic form in $A_1, \ldots, A_k$, small values of SM represent agreement with H_0 (7.2). When the τ's are not all equal, we would expect a portion of the A_j's to differ from their common null expectation of zero, with some tending to be positive and some tending to be negative. The net effect would be a large value of the quadratic form SM. This quite naturally suggests rejecting H_0 in favor of H_1 (7.3) for large values of SM and motivates procedures (7.53) and (7.54).

60. *Special Cases.* When the configuration of observed data in each of the blocks satisfies the constraints of a BIBD (see Section 7.6), the procedures in (7.53) and (7.54) are equivalent to the exact and large-sample approximation forms, respectively, of the Durbin–Skillings–Mack test procedure given in (7.44) and (7.45), respectively Section 7.6. Moreover, when we have an observation present in every treatment–block combination (i.e., we have a randomized complete block design), the Skillings–Mack procedures in (7.53) and (7.54) are equivalent to the exact and large-sample approximation forms, respectively, of the Friedman test procedure presented in (7.6) and (7.7), respectively, of Section 7.1. Thus,

the Skillings–Mack procedures in (7.53) and (7.54) represent natural extensions of the most commonly used nonparametric procedures for randomized complete block and balanced incomplete block designs to the setting of arbitrary incompleteness of the blocks.

We note that there is also an alternative closed-form expression for the test statistic *SM* (7.52) when only a single treatment has missing data. Without loss of generality, suppose that treatment k is missing observations in blocks $t+1$, $t+2,\ldots,n$ (i.e., treatment k has only t observations). Then, it can be shown (see Problem 76) that *SM* can be written as

$$SM = [t + (k-1)n]^{-1} \left\{ \sum_{j=1}^{k-1} A_j^2 + \left[\frac{nA_k^2}{t} \right] \right\}. \tag{7.55}$$

61. *Assumptions.* We emphasize that Assumption A3 stipulates that the *ns* cell distributions F_{ij} for those treatment–block combinations where observations are collected can differ at most in their locations (medians) and that these location differences (if any) must be a result of additive block and/or treatment effects (i.e., there is no interaction between the treatment and block factors). In particular, Assumption A3 requires that the *ns* underlying distributions belong to the same general family (F) and that they do not differ in scale parameters (variability). We do note, however, that the test procedure (7.53) remains distribution-free under the less restrictive setting where Assumption A3 is replaced by the weaker condition Assumption A3′ stated in Comment 43. Assumption A3 then corresponds to Assumption A3′ with the additional condition that $F_1 \equiv \cdots \equiv F_n$. (See also Comment 58.)

62. *Use of Generalized Inverse.* We noted in the body of the text that *any* generalized inverse $\sum_0^-$ can be used in the computation of *SM* (7.52). Skillings and Mack (1981) have shown that the value of *SM* is invariant with respect to the choice of generalized inverse, so that there is no ambiguity in the computation of *SM* and the associated test procedures (7.53) and (7.54) even if $\sum_0$ (7.50) is not of full rank. Of course, as we also noted previously, if $\lambda_{qt} > 0$ for all $q \neq t = 1,\ldots,k$, then the rank of the covariance matrix $\sum_0$ is, in fact, $k-1$ and we can simply use the ordinary inverse $\sum_0^{-1}$ in the definition of *SM*.

63. *Weighting Factor.* In the computation of the weighted sums of centered ranks $A_1,\ldots,A_k$, Skillings and Mack (1981) chose to weight the within-blocks centered ranks $[r_{ij} - (s_i+1)/2]$ by the factor $[12/(s_i+1)]^{1/2}$. They noted that this weighting factor has several advantages over other alternatives. First, it leads to a simple null covariance structure $\sum_0$ (7.50), which is useful for computational purposes. Second, because the range of the $[r_{ij} - (s_i+1)/2]$ values is less for blocks having fewer observations than for complete blocks, it is quite reasonable to adjust for this fact by using larger weights in those blocks with fewer observations. This has the effect of equalizing the contribution of each block when computing the A_j's. Prentice (1979) showed that a similar weighting scheme (using weights of $[s_i+1]^{-1}$) leads to increased power over use of the unweighted forms of the treatment sums. Skillings and Mack (1981) also note that use of the alternative simple weights s_i^{-1} would minimize the null variance of the weighted sums of centered ranks $A_1,\ldots,A_k$. However, the use of these simple weights would also alter the simplicity of the null covariance matrix $\sum_0$,

and computation of the associated test statistic would be much more difficult than is the case for *SM* (7.52). Other weighting schemes have been considered by Benard and van Elteren (1953) and Brunden and Mohberg (1976).

64. *Derivation of the Distribution of SM under H_0 (No-Ties Case).* The null distribution of *SM* (7.52) can be obtained by using the fact that under H_0 (7.2), all possible $s_1!s_2!\ldots s_k!$ rank configurations for the within-blocks ranks of the *observed* data are equally likely. We would simply compute the value of *SM* for each of these $s_1!s_2!\ldots s_k!$ block rank configurations and then tabulate the collective distribution of the values. Since the specifics of generating such a null distribution for *SM* are virtually identical with those for the Durbin–Skillings–Mack statistic *D* (7.43) for balanced incomplete block designs, the reader is referred to Comment 49 for illustration of the details of the process.

For a given arbitrary incomplete block design with incidence matrix **obs.mat**, the R command `cSkilMack(`α`, obs.mat)` can be used to find the available upper-tail critical values sm_α for possible values of *SM*. The incidence matrix is an $n \times k$ matrix of ones and zeroes, which indicate where the data are observed and unobserved, respectively. For a given available significance level α, the critical value sm_α then corresponds to $P_0(SM \geq sm_\alpha) = \alpha$ and is given by `cSkilMack(`α`, obs.mat)` = sm_α. Thus, for example, for $k = 3$, $n = 5$, all $c_{ij} = 1$ except for $c_{23} = c_{41} = c_{52} = 0$, the incidence matris is given by

$$\textbf{obs.mat} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

and we find `cSkilMack(.0208, obs.mat)` = $sm_{.0208}$ = 6.6347, so that $P_0(SM \geq 6.6347) = .0208$ for this missing data configuration.

65. *Large-Sample Approximation.* Let $\boldsymbol{A}$ be the vector defined in (7.49). Since each A_j is a weighted sum of centered ranks, it is not surprising (see Skillings and Mack (l981) for more details) that a properly standardized version of $\boldsymbol{A}$ has an asymptotic (n tending to infinity) $(k-1)$-variate normal distribution with mean vector $\mathbf{0} = (0, \ldots, 0)$ and covariance matrix $\sum_0$ (7.50) when the null hypothesis H_0 is true and $\lambda_{qt} > 0$, for every $q \neq t = 1, \ldots, k$. (Note once again that $\boldsymbol{A}$ does not include A_k, because A_k can be expressed as a weighted linear combination of $A_1, \ldots, A_{k-1}$. This is the reason that the asymptotic normal distribution is $(k-1)$-variate and not k-variate.) Since the test statistic *SM* (7.52) is a quadratic form in the variables $A_1, \ldots, A_{k-1}$, it is therefore quite natural that *SM* has an asymptotic (n tending to infinity) chi-square distribution with $k-1$ degrees of freedom when the null hypothesis H_0 is true and $\lambda_{qt} > 0$, for every $q \neq t = 1, \ldots, k$.

66. *Exact Conditional Null Distribution of SM with Ties among the Observed X Values within Blocks.* To have a test with exact significance level even in the presence of tied X's within some of the blocks, we need to consider all $s_1!s_2!\ldots s_k!$ block rank configurations for the *observed* data, where now these within-blocks ranks are obtained using average ranks to break ties. As in Comment 64, it still follows that under H_0 each of the $s_1!s_2!\ldots s_k!$ observed block rank configurations (now with these tied ranks) is equally likely. For

each such configuration, the value of SM (7.52) is computed and the results are tabulated. Since the specifics of generating such a conditional null distribution for SM in the case of tied within-blocks observations are virtually identical with those for the case of tied observations with the Durbin–Skillings–Mack statistic D (7.43) for balanced incomplete block designs, the reader is referred to Comment 51 for illustration of the details of the process.

67. *Settings Where the Chi-Square Approximation Is Not Applicable.* We note that the sole condition (other than the number of blocks becoming large) for the chi-square approximation to be applicable is that each of the λ_{qt}'s, $q \neq t = 1, \ldots, k$, must be positive. In settings where at least one of the λ_{qt}'s is zero, the approximate procedure (7.54) is not applicable. For such cases, one could still use procedure (7.53) by generating the (exact or simulated) null distribution of SM (7.52) and obtaining the appropriate critical values. On the other hand, if $\lambda_{qt} = 0$ for a particular pair of treatments q and t (so that q and t never appear together in a block), then procedure (7.53) would not necessarily be effective in testing H_0 (7.2) even when τ_q and τ_t are quite different.

68. *Historical Development.* The test procedures (7.53) and (7.54) were proposed and studied by Skillings and Mack (1981). They provided some exact null distribution critical values for the special case of BIBDs (see Section 7.6) and for a second special case where we are only a single observation short of having complete block data.

Properties

1. *Asymptotic Chi-Squaredness.* See Skillings and Mack (1981).

Problems

74. In the data on percent reduction in average wind speed due to shelterbelts discussed in Problem 19, the month of November was omitted from the data in Table 7.7 because the percent reduction observation at 20 m was missing. In Table 7.16, we again present these data with the month of November included.

Table 7.16 Percent Reduction in Average Wind Speed at Dambatta, 1980/81

	Leeward Distance from Shelterbelt m				
Month	20	40	100	150	200
January	22.1	20.7	15.4	12.3	6.9
February	19.2	18.7	14.9	9.3	6.5
March	21.5	21.9	14.3	9.9	7.1
April	21.5	21.2	11.1	9.4	6.2
May	21.3	20.9	11.2	9.4	7.7
June	20.9	19.6	16.9	11.6	7.0
August	19.3	18.7	14.4	12.5	7.0
September	20.1	19.6	15.6	12.6	7.5
October	23.7	20.4	14.6	12.4	8.5
November	19.5	18.4	13.8	8.4	—

Source: J. E. Ujah and K. B. Adeoye (1984).

Table 7.17 Assembly Times (min)

	Assembly Methods			
Workers	A	B	C	D
1	3.2	4.1	3.8	4.2
2	3.1	3.9	3.4	4.0
3	4.3	3.5	4.6	4.8
4	3.5	3.6	3.9	4.0
5	3.6	4.2	3.7	3.9
6	4.5	4.7	3.7	—
7	—	4.2	3.4	—
8	4.3	4.6	4.4	4.9
9	3.5	—	3.7	3.9

Source: J. H. Skillings and G. A. Mack (1981).

Use the Skillings–Mack procedure (7.53) to test the hypothesis that there is a difference in percent reduction in average wind speed over the five leeward distances from a shelterbelt. (Note that the conclusion of the Skillings–Mack procedure applied to these data is not directional, as was the case with the decision in Problem 19 using the Page ordered alternatives procedure and the randomized complete design data without the month of November. A corresponding ordered alternatives analog to the Skillings–Mack procedure for arbitrary missing data is not available in the literature.)

75. Skillings and Mack (1981) consider the experiment evaluating four methods of assembling a product, where the blocking factor corresponds to the individual assembly workers. The data for this experiment are presented in Table 7.17, where the observations are assembly times in minutes. The missing observations are due to machinery breakdowns or employee absenteeism. Use these data to assess whether there are any differences among the assembly methods with regard to median time for assembly of the product.

76. Verify that expression (7.55) is an alternative way to compute the statistic SM (7.52) when only a single treatment has missing data.

77. Consider Table 7.18, which gives a subset of the Rounding-first-base data in Table 7.1 obtained by Woodward (1970) in his study of the best method of rounding first base to minimize the time to second base. (The missing observations might be due to injury during one of the other runs.) Use these data to assess whether there are any differences among median times to second base for these three ways of rounding first base.

78. We noted that the value of the test statistic SM (7.52) does not depend on the form of the particular generalized inverse $\sum_0^-$ used in the calculation. Illustrate this fact by computing SM using two different generalized inverses for a setting where the rank of $\sum_0$ is not $k - 1$.

79. Consider Table 7.19, which gives a subset of the serum CPK activity data in Table 7.3 obtained by Goode and Meltzer (1976) in their study of the effect of isometric exercise on serum CPK levels. Use these data to assess whether there are any differences in median serum CPK activity (in mU/l) for the three measurement periods.

80. Verify for the data in Table 7.15 of Example 7.8 that the value of $SM = 13.287$ would also be obtained if we take $\boldsymbol{A}$ to be (A_1, A_3) or (A_2, A_3) and make the corresponding changes in the definition of the null covariance matrix $\sum_0$ (7.50).

81. Verify that the Skillings–Mack statistic SM (7.52) simplifies to the closed-form expression for the Friedman statistic S (7.5) when we have data from a randomized complete block design.

82. Verify that the Skillings–Mack statistic SM (7.52) simplifies to the closed-form expression for the Durbin–Skillings–Mack statistic D (7.43) when we have data from a balanced incomplete block design.

Table 7.18 Rounding-First-Base Times

Players	Methods		
	Round out	Narrow angle	Wide angle
1	5.40	5.50	5.55
2	5.85	5.70	5.75
3	5.20	5.60	5.50
4	5.55	5.50	—
5	5.90	5.85	5.70
6	5.45	5.55	5.60
7	5.40	5.40	5.35
8	—	5.50	5.35
9	5.25	5.15	5.00
10	5.85	—	5.70
11	5.25	5.20	5.10
12	5.65	5.55	—
13	5.60	5.35	5.45
14	5.05	—	4.95
15	5.50	5.50	5.40
16	—	5.55	5.50
17	5.55	5.55	—
18	5.45	5.50	5.55
19	5.50	5.45	5.25
20	5.65	5.60	5.40
21	5.70	5.65	5.55
22	6.30	6.30	6.25

Source: W. F. Woodward (1970).

Table 7.19 Effect of Isometric Exercise on Serum Creatine Phosphokinase (CPK) Activity (mU/l) in Psychotic Patients

Subject	Preexercise	19-h Postexercise	42-h Postexercise
1	27	101	82
2	30	112	50
3	24	26	68
4	54	89	—
5	21	30	49
6	36	41	48
7	36	29	46
8	16	20	8
9	21	26	25

Source: D. J. Goode and H. Y. Meltzer (1976).

83. Consider the setting corresponding to $k = 4$ and $n = 10$ where we have only a single missing observation in one of the blocks. Compare the critical region for the exact level $\alpha = .0501$ test of H_0 (7.2) based on SM with the critical region for the corresponding nominal level $\alpha = .0501$ test based on the large-sample approximation.

84. Consider the setting corresponding to $k = 6$ and $n = 5$ where we have only a single missing observation in one of the blocks. Compare the critical region for the exact level $\alpha = .0499$ test of H_0 (7.2) based on SM with the critical region for the corresponding nominal level $\alpha = .0499$ test based on the large-sample approximation.

85. Consider the incomplete block data setting corresponding to $k = 3$, $n = 3$, $s_1 = s_2 = 3$, and $s_3 = 2$. Obtain the form of the exact null H_0 distribution of SM (7.52) for the case of no-tied observations.

86. Consider the incomplete block data setting corresponding to $k = 3$, $n = 3$, $s_1 = s_2 = 3$, and $s_3 = 2$. Suppose the three observations in block 2 are all tied at a single value, but there are no tied observations in any of the other blocks. Obtain the conditional exact probability distribution of SM (7.52) under H_0 (7.2) when average ranks are used to break this set of within-block ties. Compare this conditional null distribution of SM with the null distribution of SM obtained in Problem 85 when there are no ties.

REPLICATIONS—TWO-WAY LAYOUT WITH AT LEAST ONE OBSERVATION FOR EVERY TREATMENT–BLOCK COMBINATION

It is often the case in two-way layout settings that we have more than one observation for some of the treatment–block combinations. These multiple observations in a given cell are referred to as replications for that treatment–block combination. Of course, permitting such replications opens the possibility of a much wider variety of data configurations for our two-way layout. There could be some cells with no observations, some with one observation, and some with multiple observations. In the next two sections we emphasize nonparametric procedures for general alternatives in the setting where we have a common, equal number $c > 1$ of replications for every treatment–block combination. Direct extensions of these general alternatives procedures to less restrictive settings where the number of replications need not be equal but there are no empty cells are discussed in Comment 77. Nonparametric procedures that are valid for the most general two-way layout settings where there may be a mix of cells with more than one observation (i.e., replications), cells with a single observation, and empty cells with no observations are discussed in Comment 78 for the cases of general and ordered alternatives.

In Section 7.9 we present a distribution-free hypothesis test for general alternatives when we have an equal number (>1) of replications for every treatment–block combination. In Section 7.10 we discuss an all-treatments multiple comparison procedure for the same setting.

Throughout these two sections, we continue to operate under the general conditions of Assumptions A1–A3. However, in Sections 7.9 and 7.10, we impose the additional constraint that each c_{ij} is equal to $c(>1)$ and, thus, that

$$N = \sum_{i=1}^{n} \sum_{j=1}^{k} c_{ij} = nkc.$$

7.9 A DISTRIBUTION-FREE TEST FOR GENERAL ALTERNATIVES IN A RANDOMIZED BLOCK DESIGN WITH AN EQUAL NUMBER $C(>1)$ OF REPLICATIONS PER TREATMENT–BLOCK COMBINATION (MACK–SKILLINGS)

In this section we present a procedure for testing H_0 (7.2) against the general alternatives H_1 (7.3) for block data where we have an equal number $c > 1$ replications for each of the treatment–block combinations. Here, the total number of observations is $N = nkc$.

Procedure

To compute the Mack–Skillings statistic for this equal replications setting, we first rank the observations from least to greatest separately within each of the n blocks. Let r_{ijq} be the within-block rank of X_{ijq} (the qth replication from the jth treatment in the ith block) among the kc total observations present in the ith block, for $i = 1, \ldots, n$. Set

$$S_j = \sum_{i=1}^{n} \left[\sum_{q=1}^{c} r_{ijq}/c \right], \quad \text{for } j = 1, \ldots, k. \tag{7.56}$$

Thus, S_j is the sum (across blocks) of the cellwise averages of the within-blocks ranks assigned to the c observations from treatment j, for $j = 1, \ldots, k$. The Mack–Skillings statistic for equal replications is then given by

$$\begin{aligned} MS &= \left[\frac{12}{k(N+n)} \right] \sum_{j=1}^{k} \left[S_j - \frac{N+n}{2} \right]^2, \\ &= \left[\frac{12}{k(N+n)} \right] \left\{ \sum_{j=1}^{k} S_j^2 \right\} - 3(N+n), \end{aligned} \tag{7.57}$$

where $(N+n)/2n = (kc+1)/2 = \sum_{j=1}^{k} \sum_{q=1}^{c} r_{ijq}/kc$ is the average within-blocks rank assigned for each of the n blocks. It follows that $n(N+n)/2n = (N+n)/2$ is the expected sum (across blocks) of the cellwise averages for each of the k treatments when H_0 (7.2) is true; that is, $(N+n)/2$ is the expected value of S_j, for each $j = 1, \ldots, k$, when the null hypothesis H_0 is true.

To test

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } MS \geq ms_\alpha; \quad \text{otherwise do not reject,} \tag{7.58}$$

where the constant ms_α is chosen to make the type I error probability equal to α. The constant ms_α is the upper α percentile for the null $(\tau_1 = \cdots = \tau_k)$ distribution of MS. Comment 73 explains how to obtain the critical values ms_α for k treatments, n blocks, c replications for each treatment–block combination, and available values of α.

Large-Sample Approximation

When H_0 (7.2) is true, the statistic MS has, as the common number of observations on each treatment, nc, tends to infinity, an asymptotic chi-square (χ^2) distribution with $k-1$ degrees of freedom. (See Comment 74 for indications of the proof.) The chi-square

approximation for procedure (7.58) is

$$\text{Reject } H_0 \text{ if } MS \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.59}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile of a chi-square distribution with $k-1$ degrees of freedom. To find $\chi^2_{k-1,\alpha}$, we use the R command `qchisq`$(1-\alpha, k-1)$. For example, to find $\chi^2_{6,.025}$, we apply `qchisq(.975, 6)` and obtain $\chi^2_{6,.025} = 14.45$.

Mack and Skillings (1980) have pointed out that this chi-square approximation is adequate when the significance level α is at least .05 and the number of replications c is at least 4, even though it is slightly conservative when the level nears .05. However, for significance levels as low as .01, they note that the conservative nature of the approximate procedure (7.59) can be somewhat severe unless the common number of replications c is rather large. Whenever possible, they recommend the use of the exact procedure (7.58) for such small significance levels.

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual sums of cellwise averages of the within-blocks ranks $S_1, \ldots, S_k$. In such cases, the significance level associated with procedure (7.58) is only approximately equal to α. (See Comment 75 for discussion of how to construct an exact conditionally distribution-free test of H_0 even when there are tied observations within some of the blocks.)

EXAMPLE 7.9 ***Determination of Niacin in Bran Flakes.***

In a study to investigate the precision and homogeneity of a procedure for assessing the amount of niacin in bran flakes, Campbell and Pelletier (1962) prepared homogenized samples of bran flakes enriched with 0, 4, or 8 mg niacin per 100 g of cereal. Portions of the homogenized samples were sent to different laboratories, which were asked to carry

Table 7.20 Amount of Niacin in Enriched Bran Flakes

	Amount of niacin enrichment (milligrams per 100 g bran flakes)		
Laboratory	0	4	8
1	7.58 (3)	11.63 (7)	15.00 (2)
	7.87 (8)	11.87 (11)	15.92 (9)
	7.71 (6)	11.40 (3)	15.58 (4)
2	8.00 (9.5)	12.20 (12)	16.60 (12)
	8.27 (12)	11.70 (8.5)	16.40 (11)
	8.00 (9.5)	11.80 (10)	15.90 (7)
3	7.60 (4)	11.04 (2)	15.87 (6)
	7.30 (1)	11.50 (5.5)	15.91 (8)
	7.82 (7)	11.49 (4)	16.28 (10)
4	8.03 (11)	11.50 (5.5)	15.10 (3)
	7.35 (2)	10.10 (1)	14.80 (1)
	7.66 (5)	11.70 (8.5)	15.70 (5)

Source: J. A. Campbell and O. Pelletier (1962).

out the specified procedure for each of three separate samples. The resulting data (in milligrams per 100 g bran flakes) for a subset (4 out of 12) of the laboratories included in the study are presented in Table 7.20.

Of primary interest here is the precision of the laboratory procedure for determining niacin content in bran flakes. The actual amount of niacin enrichment in the prepared bran flakes serves only as a "nuisance" blocking factor in our evaluation of the consistency of the results across the four laboratories for which data are included in Table 7.20. Hence, we have data from a two-way layout with $k = 4$ treatments (laboratories), $n = 3$ blocks (amounts of niacin enrichment), and $c = 3$ replications (individual bran flake samples) per laboratory/enrichment combination. For the purpose of illustration, we consider the significance level $\alpha = .0501$. Applying the R command `cMackSkil(`α`,k,n,c)` with $k = 4$, $n = 3$, and $c = 3$, we see that `cMackSkil(.0501,4,3,3)` $= ms_{.0501} = 7.479$ and procedure (7.58) becomes

$$\text{Reject } H_0 \text{ if } MS \geq 7.479.$$

Now, we illustrate the computations leading to the sample value of MS. The numbers in parentheses after the data values in Table 7.20 are the within-enrichment-levels (i.e., blocks) ranks (using average ranks to break ties) of the niacin content measurements obtained from the four laboratories. Using these block ranks, we obtain the following sums of cellwise averages for the four laboratories:

$$S_1 = \frac{3+8+6+7+11+3+2+9+4}{3} = 17.67,$$

$$S_2 = \frac{9.5+12+9.5+12+8.5+10+12+11+7}{3} = 30.5,$$

$$S_3 = \frac{4+1+7+2+5.5+4+6+8+10}{3} = 15.83,$$

and

$$S_4 = \frac{11+2+5+5.5+1+8.5+3+1+5}{3} = 14.$$

Hence, with $k = 4$, $n = 3$, and $N = 36$, we find from (7.57) that

$$MS = \left[\frac{12}{4(36+3)}\right]\{(17.67)^2 + (30.5)^2 + (15.83)^2 + (14)^2\} - 3(36+3)$$

$$= \left[\frac{1}{13}\right]\{312.23 + 930.25 + 250.59 + 196\} - 117 = 12.93.$$

Since the observed value of MS is greater than the critical value 7.479, we can reject H_0 at the $\alpha = .0501$ level, providing rather strong evidence that the studied process for assessing niacin content in bran flakes does not produce consistent results across a variety of laboratories and is therefore not reliable as an evaluative procedure. In fact, from the observed value of $MS = 12.93$, we can use the R command `pMackSkil(`**niacin**`)` to find that $P_0(MS \geq 12.93) =$ `pMackSkil(`**niacin**`)` $= .0023$. Thus, the smallest significance level at which we can reject H_0 in favor of H_1 with the observed value of the test statistic $MS = 12.93$ is .0023.

We should note in passing that there also appears to be an even more basic problem with this studied procedure for assessing niacin content in bran flakes and that is

the accuracy (in addition to the lack of reproducibility) of the numerical values of the measurements. For example, for those samples enriched with 4 mg niacin per 100 g bran flakes, the values obtained by applying this procedure to the sample bran flakes ranged from 10.10 to 12.20 mg per 100 g bran flakes, well over the preestablished niacin content. (Similar comments apply to the 0- and 8-mg enrichment samples.) This clearly indicates a rather severe basic calibration problem with the assessment procedure, in addition to the lack of portability across laboratories detected by our application of the Mack–Skillings procedure to these data.

Comments

69. *More General Setting.* We could replace Assumptions A1–A3 and H_0 (7.2) with the more general null hypothesis that all possible $[(ck)!]^n$ configurations for the permutations of the within-blocks ranks (r_{ijq}'s) are equally likely. Procedure (7.58) remains distribution-free for this more general null hypothesis.

70. *Motivation for the Test.* Under Assumptions A1–A3 and H_0 (7.2), each of the block rank vectors $\boldsymbol{R}_i^* = (r_{i11}, \ldots, r_{i1c}, r_{i21}, \ldots, r_{i2c}, \ldots, r_{ik1}, \ldots, r_{ikc})$, $i = 1, \ldots, n$, has a uniform distribution over the set of all $(ck)!$ permutations of the vector of integers $(1, 2, \ldots, ck)$ and this is true, independently, for each of the n blocks. It is then the case that $E_0(r_{ijq}) = (ck + 1)/2$ for every $i = 1, \ldots, n$; $j = 1, \ldots, k$; and $q = 1, \ldots, c$. It follows from (7.56) that $E_0(S_j) = nc(ck + 1)/2c = (nck + n)/2 = (N + n)/2$. Since the test statistic MS (7.57) is a constant times a sum of squared differences between the observed treatment sums of cellwise average ranks, S_j, and their common null expected value, $E_0(S_j) = (N + n)/2$, small values of MS represent agreement with H_0 (7.2). When the τ's are not all equal, we would expect a portion of the associated treatment sums of cellwise average ranks, S_j, to differ from their common null expectation, $(N + n)/2$, with some tending to be smaller and some larger. The net result (after squaring the observed differences to obtain the $[S_j - (N + n)/2]^2$ terms) would be a large value of MS. This quite naturally suggests rejecting H_0 in favor of H_1 (7.3) for larger values of MS and motivates procedures (7.58) and (7.59).

71. *Special Case of $c = 1$.* When we have a single observation for every treatment–block combination (i.e., $c = 1$), we are dealing with data from a complete randomized block design. In this setting, the Mack–Skillings statistic MS (7.57) is equivalent to the Friedman statistic S (7.5). Thus, the Mack–Skillings procedures (7.58) and (7.59) represent natural extensions of the Friedman procedures (7.6) and (7.7), respectively, to the case of an equal number $c > 1$ of replications per cell.

72. *Assumptions.* We emphasize that Assumption A3 stipulates that the nk cell distributions F_{ij} can differ at most in their locations (medians) and that these location differences (if any) must be a result of additive block and/or treatment effects (i.e., there is no interaction between the treatment and block factors). In particular, Assumption A3 requires that the ns underlying distributions belong to the same general family (F) and that they do not differ in scale parameters (variability). We do note, however, that the test procedure (7.58) remains distribution-free under the less restrictive setting where Assumption A3 is replaced by the weaker condition Assumption A3′ stated in Comment 43.

Assumption A3 then corresponds to Assumption A3′ with the additional condition that $F_1 \equiv \cdots \equiv F_n$. (Also see Comment 69.)

73. *Derivation of the Distribution of MS under H_0 (No-Ties Case).* The null distribution of *MS* (7.57) can be obtained by using the fact that under H_0 (7.2), all possible $[(ck)!]^n$ configurations for the permutations of the within-blocks ranks (r_{ijq}'s) are equally likely. Thus, to obtain the exact null distribution of *MS*, we compute its value for each of these $[(ck)!]^n$ block rank configurations and then tabulate the collected outcomes. We must point out, of course, that the number $[(ck)!]^n$ of configurations for which we need to compute the value of *MS* can get large rather quickly, as either k or c is moderately increased. Since the specifics of generating such a null distribution for *MS* are virtually identical with those for the Durbin–Skillings–Mack statistic *D* (7.43) for balanced incomplete block designs, the reader is referred to Comment 49 for illustration of the details of the process.

For a given number of treatments k, blocks n, and c replications for each treatment–block combination, the R command `cMackSkil(`α`,k,n,c)` can be used to find the available upper-tail critical values ms_α for possible values of *MS*. For a given available significance level α, the critical value ms_α then corresponds to $P_0(MS \geq ms_\alpha) = \alpha$ and is given by `cMackSkil(`α`,k,n,c)` $= ms_\alpha$. Thus, for example, for $k = 4$, $n = 4$, and $c = 3$, we have $P_0(MS \geq 7.667) = .0502$, so that $ms_{.0502} =$ `cMackSkil(.0502,4,4,3)` $= 7.667$ for $k = 4$, $n = 4$, and $c = 3$.

74. *Large-Sample Approximation.* Define the centered treatment sums of cellwise average ranks $S_j^* = S_j - E_0(S_j) = S_j - (N+n)/2$, for $j = 1, \ldots, k$, and set $\boldsymbol{S}^* = (S_1^*, \ldots, S_{k-1}^*)$. Since each S_j is an average, it is not surprising (see Mack and Skillings (1980) for more details) that a properly standardized version of $\boldsymbol{S}^*$ has an asymptotic (*nc* tending to infinity) $(k-1)$-variate normal distribution with mean vector $\mathbf{0} = (0, \ldots, 0)$ and appropriate covariance matrix Σ^* when the null hypothesis H_0 is true. (Note that $\boldsymbol{S}^*$ does not include S_k^*, because S_k^* can be expressed as a weighted linear combination of $S_1^*, \ldots, S_{k-1}^*$. This is the reason that the asymptotic normal distribution is $(k-1)$-variate and not k-variate.) Since the test statistic *MS* (7.57) is a quadratic form in the variables (S_1^*, …, S_{k-1}^*), it is therefore quite natural that *MS* has an asymptotic (*nc* tending to infinity) chi-square distribution with $k-1$ degrees of freedom when the null hypothesis H_0 is true.

75. *Exact Conditional Null Distribution of MS with Ties among the X Values within Blocks.* To have a test with exact significance level even in the presence of tied *X*'s within some of the blocks, we need to consider all $[(ck)!]^n$ block rank configurations for the observed data, where now these within-blocks ranks are obtained using average ranks to break ties. As in Comment 73, it still follows that under H_0 each of the $[(ck)!]^n$ observed block rank configurations (now with these tied ranks) is equally likely. For each such configuration, the value of *MS* (7.57) is computed and the results are tabulated. Since the specifics of generating such a conditional null distribution for *MS* in the case of tied within-blocks observations are virtually identical with those for the case of tied observations with the Durbin–Skillings–Mack statistic *D* (7.43) for balanced incomplete block designs, the reader is referred to Comment 51 for illustration of the details of the process.

76. *Simple Competitor Procedure when the Number of Replications Is the Same for Every Cell.* As an alternative to the Mack–Skillings procedure (7.58), we could first compute the median of the c replications separately in each of the nk cells and then apply either the Friedman procedure (7.6) or the Page procedure (7.11), whichever is appropriate for the alternatives of interest, to these nk cell medians (which now represent data from a complete randomized block design). In general, this approach could result in substantial loss of information, especially when the number of replications per cell, c, is large. However, it is simple and does provide the only available nonparametric procedure for dealing specifically with ordered alternatives when we have an equal number (>1) of replications per cell. (We note, in passing, that any appropriate measure of central tendency, such as the cell means or the medians of the Walsh averages (see Comment 3.17), for the individual cell data, could be used instead of the cell medians to summarize the data prior to application of the Friedman or the Page procedure.)

77. *Extension to Arbitrary Replication* (≥ 1) *Configurations.* We have described the Mack–Skillings procedure in detail for the setting where we have the same number of replications c (≥ 1) for each of the treatment–block combinations. However, in their original work, Mack and Skillings (1980) proposed a more general test procedure that is appropriate for any two-way layout setting for which we have at least one replication for every treatment–block combination (i.e., there are no empty cells). We now present their procedure for this more general setting where the only stipulation is that $c_{ij} > 0$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

For $i = 1, \ldots, n$, let $q_i = \sum_{j=1}^{k} c_{ij}$ be the total number of observations present in the ith block. Once again we rank the observations from least to greatest within each of the blocks and let r_{iju} denote the rank of X_{iju} within the q_i observations present in the ith block, for $u = 1, \ldots, c_{ij}$; $i = 1, \ldots, n$; and $j = 1, \ldots, k$. For each treatment, compute the sum of cellwise weighted average ranks

$$V_j = \sum_{i=1}^{n} \sum_{u=1}^{c_{ij}} \frac{r_{iju}}{q_i}, \quad j = 1, \ldots, k. \tag{7.60}$$

Define the vector

$$\begin{aligned} \boldsymbol{V} &= (V_1 - E_0[V_1], \ldots, V_{k-1} - E_0[V_{k-1}]) \\ &= \left(V_1 - \sum_{i=1}^{n} \left[\frac{c_{i1}(q_i + 1)}{2q_i} \right], \ldots, V_{k-1} - \sum_{i=1}^{n} \left[\frac{c_{i,k-1}(q_i + 1)}{2q_i} \right] \right). \end{aligned} \tag{7.61}$$

Thus, the components of $\boldsymbol{V}$ are the sums of cellwise weighted average ranks centered about their expected values under H_0. (Without the loss of generality, we have chosen to omit the centered V_k from the vector $\boldsymbol{V}$. The V_j's are linearly dependent, because a weighted linear combination of all k of them is a constant. We could omit any one of the V_j's in the definition of $\boldsymbol{V}$ and the procedure we now describe would lead to the same value of the test statistic. For further discussion, see Mack and Skillings (1980).)

The covariance matrix for $\boldsymbol{V}$ under H_0 (7.2) has the form $\sum_{\boldsymbol{V},\mathbf{0}} = ((\sigma_{s,t}))$, where

$$\sigma_{s,t} = \sum_{i=1}^{n} \left[\frac{c_{is}(q_i - c_{is})(q_i + 1)}{12q_i^2} \right], \quad \text{for } s = t = 1, \ldots, k-1$$

$$= -\sum_{i=1}^{n} \left[\frac{c_{is}c_{it}(q_i + 1)}{12q_i^2} \right], \quad \text{for } s \neq t = 1, \ldots, k-1. \tag{7.62}$$

The rank of the matrix $\boldsymbol{\Sigma}_{\boldsymbol{V},\mathbf{0}}$ is $k-1$. Letting $\boldsymbol{\Sigma}_{\boldsymbol{V},\mathbf{0}}^{-1}$ denote the inverse of $\boldsymbol{\Sigma}_{\boldsymbol{V},\mathbf{0}}$, the Mack–Skillings test statistic for this general setting of unequal, but positive, numbers of replications in the treatment–block combinations, is given by

$$MS_g = \boldsymbol{V}\,\boldsymbol{\Sigma}_{\boldsymbol{V},\mathbf{0}}^{-1}\boldsymbol{V}'. \tag{7.63}$$

To test

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \tau_2, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance, the Mack–Skillings general procedure is then to

$$\text{Reject } H_0 \text{ if } MS_g \geq ms_{g,\alpha}; \quad \text{otherwise do not reject,} \tag{7.64}$$

where the constant $ms_{g,\alpha}$ is chosen to make the type I error probability equal to α.

The critical values $ms_{g,\alpha}$ are available in the literature only for the setting where we have an equal number, c, of replications in each cell, in which case the general Mack–Skillings test procedure (7.64) is equivalent to the equal replications version given in (7.58). However, when H_0 (7.2) is true, the general form statistic MS_g, has, as N tends to infinity in such a way that c_{ij}/N tends to $\rho_{ij} > 0$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, k$, an asymptotic chi-square (χ^2) distribution with $k-1$ degrees of freedom. Thus, when N is large, the chi-square approximation for the general Mack–Skillings procedure (7.64) is

$$\text{Reject } H_0 \text{ if } MS_g \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.65}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile of a chi-square distribution with $k-1$ degrees of freedom.

78. *Competitor Procedures Applicable for Most General Two-Way Layout Settings Where There Are Both Replications and Empty Cells.* Thus far in this chapter we have discussed procedures that are appropriate either for settings where we have 0 or 1 observation for every treatment–block combination or for settings where we have at least one observation in every cell. None of these procedures are appropriate for the most general settings that represent a combination of these two structures, namely, those data sets where we have replications ($c_{ij} > 1$) for

some treatment–block combinations and no observations ($c_{ij} = 0$) for others. We briefly discuss now two test procedures for such general two-way layout settings, one designed for general alternatives to H_0 (7.2) and the second specifically oriented toward detecting ordered alternatives.

Let k_i be the number of treatments in the ith block for which $c_{ij} > 0$, for $i = 1, \ldots, n$. (Once again, we discard any block i for which $k_i = 1$, as such a block contains no information relative to possible differences in the treatment effects. Notationally, then, n represents the number of blocks remaining after discarding blocks with observations on only a single treatment.)

General Alternatives. We first compute the one-way layout Kruskal–Wallis statistic H (6.5) separately in each of the n blocks. Letting H_i denote this Kruskal–Wallis statistic for the ith block, $i = 1, \ldots, n$, the statistic considered by Mack (1981) for this most general two-way layout setting is given by

$$H_{\text{tot}} = \sum_{i=1}^{n} H_i. \tag{7.66}$$

The level α test of H_0 (7.2) versus the general alternatives H_1 (7.3) studied by Mack (1981) is

$$\text{Reject } H_0 \text{ if } H_{\text{tot}} \geq h_\alpha^*; \quad \text{otherwise do not reject,} \tag{7.67}$$

where the constant h_α^* is chosen to make the type I error probability equal to α. Values of h_α^* are available in Mack (1981) for $k = 3$, $n = 2$, 3, and all combinations of replications $0 \leq c_{ij} \leq 3$, as well as for $k = 3, n = 4$, 5, and all combinations of replications $0 \leq c_{ij} \leq 2$. Additional values of h_α^* can be found in DeKroon and Van der Laan (1981) for $\alpha = .01$, .05, and various combinations of k, n, and equal number of replications c in the ranges $2 \leq k \leq 4$, $1 \leq n \leq 10$, and $2 \leq c \leq 4$. (We note, in passing, that procedure (7.67) can be particularly sensitive to a large degree of interaction between the treatment and the block factors. In the presence of such extensive interaction, it is possible that a rejection of H_0 with procedure (7.67) could be a direct consequence of this interaction, rather than because of any significant differences in the treatment effects $\tau_1, \ldots, \tau_k$.)

When H_0 (7.2) is true, the statistic H_{tot} has, as min (nonzero c_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, k$) tends to infinity, an asymptotic chi-square (χ^2) distribution with $d = (k_1 + k_2 + \cdots + k_n - n)$ degrees of freedom (see Mack (1981) for details). Thus, when the minimum nonzero c_{ij} is large, the chi-square approximation for procedure (7.67) is

$$\text{Reject } H_0 \text{ if } H_{\text{tot}} \geq \chi^2_{d,\alpha}; \quad \text{otherwise do not reject,} \tag{7.68}$$

where $\chi^2_{d,\alpha}$ is the upper α percentile point of a chi-square distribution with d degrees of freedom.

Ordered Alternatives. If we are interested in ordered alternatives, H_2 (7.9), we first compute the one-way layout Jonckheere–Terpstra statistic J (6.13) separately in each of the n blocks. Letting J_i denote this Jonckheere–Terpstra statistic for the ith block, $i = 1, \ldots, n$, the statistic proposed by Skillings and

Wolfe (1977, 1978) for this most general two-way layout ordered alternatives setting is given by

$$J_{\text{tot}} = \sum_{i=1}^{n} J_i. \tag{7.69}$$

The level α test of H_0 (7.2) versus the ordered alternatives H_2 (7.9) suggested by Skillings and Wolfe (1977, 1978) is

$$\text{Reject } H_0 \text{ if } J_{\text{tot}} \geq j_\alpha^*; \quad \text{otherwise do not reject,} \tag{7.70}$$

where the constant j_α^* is chosen to make the type I error probability equal to α. Values of j_α^* are available in Skillings (1980) for $k = 2(1)6$, $n = 2(1)5$ and selected configurations of the c_{ij}'s such that $c_{ij} = C_i$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$ (i.e., within a given block, each treatment has the same number of observations C_i, but C_1, C_2, ..., C_n need not all be equal). (We note that procedure (7.70) does not have the same sensitivity to the presence of extensive interaction as does the general alternatives procedure (7.67). Rejection of H_0 with procedure (7.70) will always be indicative of the presence of an ordered structure on the treatment effects $\tau_1, \ldots, \tau_k$.)

When H_0 (7.2) is true, the standardized form

$$J_{\text{tot}}^* = \frac{J_{\text{tot}} - E_0(J_{\text{tot}})}{[\text{var}_0(J_{\text{tot}})]^{1/2}} \tag{7.71}$$

has, as min (nonzero $c_{ij}, i = 1, \ldots, n; j = 1, \ldots, k$) tends to infinity, an asymptotic $N(0, 1)$ distribution (see Skillings and Wolfe (1977, 1978) for details), where

$$E_0(J_{\text{tot}}) = \frac{\sum_{i=1}^{n} \left[q_i^2 - \sum_{j=1}^{k} c_{ij}^2 \right]}{4} \tag{7.72}$$

and

$$\text{var}_0(J_{\text{tot}}) = \frac{\sum_{i=1}^{n} \left[q_i^2(2q_i + 3) - \sum_{j=1}^{k} c_{ij}^2(2c_{ij} + 3) \right]}{72}, \tag{7.73}$$

are the expected value and variance, respectively, of J_{tot} (7.69) under the null hypothesis H_0 and $q_i = c_{i1} + \cdots + c_{ik}$ is the total number of observations present in the ith block, $i = 1, \ldots, n$. Thus, when the minimum nonzero c_{ij} is large, the normal theory approximation for procedure (7.70) is

$$\text{Reject } H_0 \text{ if } J_{\text{tot}}^* \geq z_\alpha; \quad \text{otherwise do not reject.} \tag{7.74}$$

79. *Historical Development.* Mack and Skillings (1980) proposed and studied a general test procedure for an arbitrary two-way layout setting where we have at least one observation for every treatment–block combination (see Comment 77). For the special case of an equal number of replications, c, in every cell, their general test procedure simplifies to the expression in (7.58) based on the test statistic *MS*. They also provided some exact null distribution critical values sm_α in this equal replications setting for a variety of combinations of k, n, and c.

Properties

1. *Asymptotic Chi-Squaredness*. See Mack and Skillings (1980).
2. *Efficiency.* See Mack and Skillings (1980) and Section 7.16.

Problems

87. Rice (1988) considered an experiment to determine whether two forms of iron, Fe^{2+} and Fe^{3+}, are retained differently, with the goal of comparing their potentials for use as dietary supplements. A total of 108 mice were randomly divided into six groups of 18 mice each. Three of these groups were given Fe^{2+} in the different concentrations, 10.2, 1.2, and .3 mM, and three groups were given Fe^{3+} in the same concentrations. The iron was radioactively labeled so that a counter could be used to accurately measure the initial amount given, and it was administered orally to the mice. At a later time, a second count was obtained on each mouse, and the percentage of iron retained was recorded. The data in Table 7.21 are the percentages retained by each of the 108 mice.

Use the Mack–Skillings large-sample procedure (7.59) to test the hypothesis that there is a difference across the concentrations studied between the two forms of iron Fe^{2+} and Fe^{3+} in percentage iron retained.

88. Let V_j be as defined in expression (7.60), for $j = 1, \ldots, k$. Show that

$$E_0[V_j] = \sum_{i=1}^{n} \left[\frac{c_{ij}(q_i + 1)}{2q_i} \right],$$

as noted in expression (7.61), where q_i is the number of observations present in the ith block, for $i = 1, \ldots, n$.

Table 7.21 Percentage of Iron Retained

Concentration	Form of iron					
	Fe^{2+}			Fe^{3+}		
.3 millimolar	2.71	5.43	6.38	2.25	3.93	5.08
	6.38	8.32	9.04	5.82	5.84	6.89
	9.56	10.01	10.08	8.50	8.56	9.44
	10.62	13.80	15.99	10.52	13.46	13.57
	17.90	18.25	19.32	14.76	16.41	16.96
	19.87	21.60	22.25	17.56	22.82	29.13
1.2 millimolar	4.04	4.16	4.42	2.20	2.93	3.08
	4.93	5.49	5.77	3.49	4.11	4.95
	5.86	6.28	6.97	5.16	5.54	5.68
	7.06	7.78	9.23	6.25	7.25	7.90
	9.34	9.91	13.46	8.85	11.96	15.54
	18.40	23.89	26.39	15.89	18.30	18.59
10.2 millimolar	2.20	2.69	3.54	0.71	1.66	2.01
	3.75	3.83	4.08	2.16	2.42	2.42
	4.27	4.53	5.32	2.56	2.60	3.31
	6.18	6.22	6.33	3.64	3.74	3.74
	6.97	6.97	7.52	4.39	4.50	5.07
	8.36	11.65	12.45	5.26	8.15	8.24

Source: J. A. Rice (1988).

89. Show that for the special case of one replication per cell (i.e., $c = 1$), the Mack–Skillings procedures (7.58) and (7.59) are equivalent to the Friedman procedures (7.6) and (7.7), respectively. (See Comment 71.)

90. Anderson and McLean (1974) considered the data from an experiment measuring the strength of a weld in steel bars. The two factors of interest in the experiment were the total time of the automatic weld cycle and the distance the weld die travels during the automatic weld cycle. Two weld-strength observations were collected at each combination of five different weld cycle times and three different weld die travel distances (gage bar settings). These weld-strength data are given in Table 7.22.

Use the Mack–Skillings procedure to test the hypothesis that weld cycle time has an effect on the strength of a weld, at least over the weld die travel distances considered in the study.

91. For the weld-strength data in Table 7.22, compute the median of the two observations in each of the gage bar setting/weld cycle time combinations. Apply the Friedman procedure (7.6) to the resulting medians to test the hypothesis that weld cycle time has an effect on the strength of a weld, at least over the weld die travel distances in the study. Compare with the result obtained in Problem 90. (See also Comment 76.)

92. Consider the Mack–Skillings statistic MS_g (7.63) for the most general two-way layout setting with at least one replication for every treatment–block combination, as discussed in Comment 77. Show that the test procedure (7.64) based on MS_g is equivalent to the equal replications test procedure (7.58) based on MS (7.57) when, in fact, we have an equal number, c, of replications for every treatment–block combination.

93. One method for the determination of coal acidity is based on the use of ethanolic NaOH. In an effort to assess the effect of the ethanolic NaOH concentration on the obtained acidity values, Sternhell (1958) studied three different NaOH concentrations (.404N, .626N, and .786N) in conjunction with three different types of coal (Morwell, Yallourn, and Maddingley). The data in Table 7.23 are the resulting acidity values determined under each of these three concentration levels for two different samples from each type of coal.

Use the Mack–Skillings procedure to test the hypothesis that the NaOH concentration has an effect on the measured coal acidity values, at least over the three types of coal included in this study.

94. Consider the percentage retained iron data in Table 7.21. Test the hypothesis that the iron concentration affects the percentage iron retention, regardless of which form of iron is involved.

Table 7.22 Strength of Weld

	Weld cycle times				
Gage bar setting	1	2	3	4	5
1	10 12	13 17	21 30	18 16	17 21
2	15 19	14 12	30 38	15 11	14 12
3	10 8	12 9	10 5	14 15	19 11

Source: V. L. Anderson and R. A. McLean (1974).

Table 7.23 Coat Acidity Value

	NaOH concentration					
Type of coal	.404N		.626N		.786N	
Morwell	8.27	8.17	8.03	8.21	8.60	8.20
Yallourn	8.66	8.61	8.42	8.58	8.61	8.76
Maddingley	8.14	7.96	8.02	7.89	8.13	8.07

Source: S. Sternhell (1958).

95. What is the maximum value for the Mack–Skillings statistic MS (7.57) when there are c replications per cell? For what rank configuration is this maximum achieved?

96. Consider the setting corresponding to $k = 4$, $n = 5$, and $c = 3$ replications per cell. Compare the critical region for the exact level $\alpha = .0100$ test of H_0 (7.2) based on MS with the critical region for the corresponding nominal level $\alpha = .0100$ test based on the large-sample approximation.

97. Consider the setting corresponding to $k = 2$, $n = 2$, and $c = 2$ replications per cell. Obtain the form of the exact null H_0 distribution of MS (7.57) for the case of no-tied observations.

98. Consider the setting corresponding to $k = 2$, $n = 2$, and $c = 2$ replications per cell. Suppose that one of the observations in the first cell (block 1 and treatment 1) is tied in value with one of the observations in the second cell (block 1 and treatment 2). Obtain the conditional exact probability distribution of MS (7.57) under H_0 (7.2) when average ranks are used to break this within-blocks tie. Compare this conditional null distribution of MS with the null distribution of MS obtained in Problem 97 when there are no ties.

99. Consider the setting corresponding to $k = 2$, $n = 2$, and $c = 2$ replications per cell. Suppose that one of the observations in the first cell (block 1 and treatment 1) is tied in value with the other observation in the same cell. Obtain the conditional exact probability distribution of MS (7.57) under H_0 (7.2) when average ranks are used to break this within-cell tie. Compare this conditional null distribution of MS with the null distributions of MS obtained in Problems 97 and 98 when there are no ties and ties between cells, respectively.

100. Consider the setting corresponding to $k = 5$, $n = 4$, and $c = 4$ replications per cell. Compare the critical region for the exact level $\alpha = .0500$ test of H_0 (7.2) based on MS with the critical region for the corresponding nominal level $\alpha = .0500$ test based on the large-sample approximation.

101. In a study to determine the effect of light on the release of luteinizing hormone (LH), Rice (1988) compared data for male and female rats kept in constant light with similar animals exposed to a regime of 14 h of light and 10 h of darkness. Five different dosages of a luteinizing release factor (LRF) were considered in the study and the measurement obtained from the animals was the level of LH (in nanograms per milliliter of serum) in blood samples collected after exposure to one of the regimes in combination with one of the LRF dosages. We consider data for the male rats only.

Sixty male rats were randomly allocated to the various experimental settings in such a way that six rats were exposed to each of the 10 combinations of light regime and LRH dosage. The LH level data for these 60 rats are given in Table 7.24.

Table 7.24 Serum Level of LH (in Nanograms per Milliliter of Serum)

	Light regime					
LRF dosage	Constant light			14 h light/10 h dark		
0 ng (control)	72	64	78	212	27	68
	20	56	70	72	130	153
10 ng	74	82	40	32	98	148
	87	78	88	186	203	188
50 ng	130	187	133	294	306	234
	185	107	98	219	281	288
250 ng	159	167	193	515	340	348
	196	174	250	205	505	432
1250 ng	137	426	178	296	545	630
	208	196	251	418	396	227

Source: J. A. Rice (1988).

Use the Mack–Skillings large-sample procedure (7.59) to test the hypothesis that degree of exposure to light has an effect on serum levels of LH across the LRH dosages included in the study.

7.10 ASYMPTOTICALLY DISTRIBUTION-FREE TWO-SIDED ALL-TREATMENTS MULTIPLE COMPARISONS FOR A TWO-WAY LAYOUT WITH AN EQUAL NUMBER OF REPLICATIONS IN EACH TREATMENT–BLOCK COMBINATION (MACK–SKILLINGS)

In this section we present an asymptotically distribution-free multiple comparison procedure using within-blocks ranks that is designed to make two-sided decisions about individual differences between pairs of treatment effects (τ_i, τ_j), for $i < j$, for data obtained from a two-way layout design with an equal number of replications for every treatment–block combination. The multiple comparison procedure of this section would generally be applied to data from such a two-way layout with an equal number of replications *after* rejection of H_0 (7.2) with the Mack–Skillings procedure from Section 7.9. In this setting we will reach conclusions about all $k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two-sided in nature.

Procedure

Let $S_1, \ldots, S_k$ be the treatment sums of cellwise averages of within-blocks ranks given by (7.56). Calculate the $k(k-1)/2$ absolute differences $|S_u - S_v|$, $1 \le u < v \le k$.

When H_0 (7.2) is true, the $k(k-1)/2$-component vector $(S_1, \ldots, S_k)$ has, when properly standardized and as N tends to infinity, an asymptotic $(k-1)$-variate normal distribution with appropriate mean vector and covariance matrix (see Mack and Skillings (1980) for details of the proof). At an approximate experimentwise error rate of α, the Mack–Skillings two-sided all-treatments multiple comparison procedure reaches its $k(k-1)/2$ pairwise decisions, corresponding to each (τ_u, τ_v) pair, $1 \le u < v \le k$, by the criterion

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |S_u - S_v| \ge [k(N+n)/12]^{1/2} q_\alpha; \quad \text{otherwise decide } \tau_u = \tau_v, \tag{7.75}$$

where q_α is the upper αth percentile for the distribution of the range of k independent $N(0,1)$ variables. To find q_α for k treatments and a specified experimentwise error rate α, we use the R command `cRangeNor`(α, k). For example, to find $q_{.01}$ for $k = 4$ treatments, we apply `cRangeNor(.01,4)` and obtain $q_{.01} = 3.240$ for $k = 4$. (See also Comment 82.)

Ties

If there are ties among the X observations within any of the blocks, use average ranks to break the ties and compute the individual sums of cellwise averages of within-blocks ranks $S_1, \ldots, S_k$.

EXAMPLE 7.10 ***Determination of Niacin in Bran Flakes.***

For the sake of illustration, we apply procedure (7.75) to the niacin determination data discussed in Example 7.9. There we had found rather strong evidence that the studied process for assessing niacin content in bran flakes does not produce consistent results across a variety of laboratories. To determine which of the laboratories differ in median detected niacin content in the bran flakes, we consider procedure (7.75) with an approximate experimentwise error rate $\alpha \approx .025$. Using the R command `cRangeNor`(α,k) with $\alpha = .025$ and $k = 4$, we find `cRangeNor(.025,4)` $= q_{.025} = 3.984$ and procedure (7.75) reduces to

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |S_u - S_v| \geq [4(36+3)/12]^{1/2}(3.984) = 14.365.$$

Using the treatments sums of cellwise averages of within-blocks ranks obtained in Example 7.9, we find that

$$|S_2 - S_1| = |30.5 - 17.67| = 12.83, \quad |S_3 - S_1| = |15.83 - 17.67| = 1.84,$$
$$|S_4 - S_1| = |14 - 17.67| = 3.67, \quad |S_3 - S_2| = |15.83 - 30.5| = 14.67,$$
$$|S_4 - S_2| = |14 - 30.5| = 16.5, \quad |S_4 - S_3| = |14 - 15.83| = 1.83.$$

Referring these differences to the approximate critical value 14.365, we see that

$$|S_2 - S_1| = 12.83 < 14.365 \quad \Rightarrow \quad \text{decide } \tau_2 = \tau_1,$$
$$|S_3 - S_1| = 1.84 < 14.365 \quad \Rightarrow \quad \text{decide } \tau_3 = \tau_1,$$
$$|S_4 - S_1| = 3.67 < 14.365 \quad \Rightarrow \quad \text{decide } \tau_4 = \tau_1,$$
$$|S_3 - S_2| = 14.67 > 14.365 \quad \Rightarrow \quad \text{decide } \tau_3 \neq \tau_2,$$
$$|S_4 - S_2| = 16.5 > 14.365 \quad \Rightarrow \quad \text{decide } \tau_4 \neq \tau_2,$$
$$\text{and} \quad |S_4 - S_3| = 1.83 < 14.365 \quad \Rightarrow \quad \text{decide } \tau_4 = \tau_3.$$

Thus, at an approximate experimentwise error rate of .025, we see that Laboratory 2 yielded significantly different median detected niacin content than either Laboratory 3 or Laboratory 4. These multiple comparison decisions help to focus the rationale for the original rejection of H_0 (7.2) by the Mack–Skillings procedure in Example 7.9, as it now seems reasonable to question the reliability of Laboratory 2 in conducting this niacin content process.

Comments

80. *Rationale for Multiple Comparison Procedure.* The rationale behind the multiple comparison procedure of this section for data from a two-way layout design with an equal number of replications is similar to that for the two-sided multiple comparison procedures for data from a complete randomized block design. For further discussion, see Comment 24.

81. *Experimentwise Error Rate.* The use of an experimentwise error rate represents a very conservative approach to multiple comparisons. We are insisting that the probability of making correct decisions be $1-\alpha$ when the null hypothesis H_0 (7.2) of treatment equivalence is true. Thus we have a high degree of protection when H_0 is true, but we often apply such techniques when we have evidence (perhaps based on a priori information or perhaps obtained by applying the Mack–Skillings test, as in Example 7.9) that H_0 is not true. The protection under H_0 also makes it harder for the procedure to judge treatments as differing significantly when, in fact, H_0 is false, and this difficulty becomes more severe as k increases. See Comment 6.54 for additional discussion of experimentwise error rates.

82. *Conservative Procedure.* Mack and Skillings (1980) also proposed a conservative multiple comparison procedure that guarantees an upper bound on the experimentwise error rate. Let $S_1, \ldots, S_k$ be the treatment sums of cellwise averages of within-blocks ranks given by (7.56). At an experimentwise error rate *no greater* than α, the Mack–Skillings conservative two-sided all-treatments multiple comparison procedure reaches its $k(k-1)/2$ decisions through the criterion

$$\text{Decide } \tau_u \neq \tau_v \text{ if } |S_u - S_v| \geq [k(N+n)ms_\alpha/6]^{1/2};$$
$$\text{otherwise decide } \tau_u = \tau_v, \tag{7.76}$$

where ms_α is the upper α percentile for the null distribution of the Mack–Skillings statistic MS (7.57). Comment 73 describes how to obtain values of ms_α for a given number of treatments k, blocks n, and c replications for each treatment–block combination. Mack and Skillings (1980) note that although procedure (7.76) does not require a large number of blocks, it is, nevertheless, rather conservative since it is based on the projection procedure of Scheffé; that is, the true experimentwise error rate might be considerably smaller than the bound α provided by (7.76). As a result, they recommend using the approximation (7.75) whenever the number of blocks is reasonably large.

83. *Dependence on Observations from Other Noninvolved Treatments.* The all-treatments multiple comparison procedure of this section suffers from the same disadvantage as do the other two-way layout multiple comparison procedures of this chapter. The decision between treatment u and treatment v can be affected by changes only in the observations from one or more of the other $k-2$ treatments that are not directly involved.

Properties

1. *Asymptotic Multivariate Normality.* See Mack and Skillings (1980).
2. *Efficiency.* See Section 7.16.

Problems

102. Apply procedure (7.75) to the weld-strength data of Table 7.22 in Problem 90.

103. Illustrate the difficulty discussed in Comment 83 by means of a numerical example.

104. Apply procedure (7.75) to the coal acidity data of Table 7.23 in Problem 93.

105. Consider the niacin content data of Table 7.20 in Example 7.9. Find the smallest approximate experimentwise error rate at which the most significant difference(s) in median bran flake niacin content between the four laboratories would be detected by procedure (7.75).

106. Consider the weld-strength data of Table 7.22 in Problem 90. Find the smallest approximate experimentwise error rate at which procedure (7.75) would declare that weld cycle times 1 and 3 have differing effects on the strength of a weld.

107. Consider the coal acidity data of Table 7.23 in Problem 93. Find the smallest approximate experimentwise error rate at which the most significant difference(s) in effects of the NaOH concentration on the measured coal acidity value would be detected by procedure (7.75).

108. Consider the coal acidity data of Table 7.23 in Problem 93. Find the smallest approximate experimentwise error rate at which procedure (7.75) would declare that there is a difference in median coal acidity level between the Morwell and the Yallourn types of coal.

ANALYSES ASSOCIATED WITH SIGNED RANKS

The statistical procedures discussed in Sections 7.1–7.5 (for randomized block designs with a single observation on each treatment–block combination) utilize the treatment observations only through comparisons within blocks. It is this restriction to within-blocks comparisons that leads directly to many of these procedures being strictly distribution-free, even for small sample sizes. An alternative approach is to consider accessing between-blocks information via utilization of pairwise signed ranks in the construction of appropriate statistical procedures. Hypothesis test and multiple comparison procedures based on these pairwise signed ranks will no longer be exactly distribution-free for small numbers (n) of blocks and they require the use of large-sample approximations. However, improved efficiency can result in many cases from this use of between-blocks signed ranks.

In the next five sections we assume (as done in Sections 7.1–7.5) that we have data from a randomized complete block design satisfying Assumptions A1–A3 for the case of one observation per treatment–block combination, corresponding to $c_{ij} = 1$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, k$. For ease of notation in these five sections, we once again drop the third subscript on the X variables, because it is always equal to 1 in this setting.

Section 7.11 contains a conservative signed ranks test procedure directed at general alternatives for randomized block designs with a single observation on each treatment–block combination, while Section 7.12 presents the corresponding conservative signed ranks test procedure designed for ordered alternatives. The associated approximate signed ranks multiple comparison procedures are given in Sections 7.13 (all-treatments comparisons) and 7.14 (treatments-versus-control comparisons). Section 7.15 contains the contrast estimators linked to the Wilcoxon signed ranks for this setting.

7.11 A TEST BASED ON WILCOXON SIGNED RANKS FOR GENERAL ALTERNATIVES IN A RANDOMIZED COMPLETE BLOCK DESIGN (DOKSUM)

In this section we present a conservative procedure based on pairwise signed ranks for testing H_0 (7.2) against the general alternative H_1 (7.3) that at least two of the treatment effects are not equal.

Procedure

For each of the $k(k-1)/2$ pairs of treatments (u, v), with $1 \le u < v \le k$, we form the n absolute differences

$$Y_{uv}^i = |X_{iu} - X_{iv}|, \quad i = 1, \ldots, n. \tag{7.77}$$

(Note that $Y_{uv}^i = |D_{uv}^i|$, where the D_{uv}^i are the same differences given in (7.36) and used in the contrast estimator discussed in Section 7.5.) For each pair of treatments (u, v), we let R_{uv}^i be the rank of Y_{uv}^i in the ranking from least to greatest of the n values $Y_{uv}^1, \ldots, Y_{uv}^n$. To compute the Doksum (1967) statistic D, set

$$T_{uv} = \sum_{i=1}^{n} R_{uv}^i \Psi_{uv}^i \quad \text{and} \quad B_{uv} = \sum_{i=1}^{n} \Psi_{uv}^i, \tag{7.78}$$

where

$$\Psi_{uv}^i = \begin{cases} 1, & \text{if } X_{iu} < X_{iv}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.79}$$

Let

$$H_{uv} = \frac{2(T_{uv} - B_{uv})}{n(n-1)}, \quad 1 \le u < v \le k. \tag{7.80}$$

(We note that the statistics H_{uv} need be calculated directly only for $u < v$, because for $u > v$, we can use the relationship $H_{vu} = 1 - H_{uv}$.) Next, we obtain the averages

$$H_{u.} = \sum_{j=1}^{k} \frac{H_{uj}}{k}, \quad u = 1, \ldots, k, \tag{7.81}$$

where we note that $H_{uu} = 0$, for $u = 1, \ldots, k$.

The common null variance of each of the $k(k-1)/2$ differences $H_{u.} - H_{v.}$, $1 \le u < v \le k$, is given by the expression

$$\text{var}_0(H_{u.} - H_{v.}) = \frac{2n - 1 + (k-2)[24(n-2)\lambda_F + 13 - 6n]}{3kn(n-1)}, \tag{7.82}$$

with

$$\lambda_F = P_0(X_1 < X_2 + X_3 - X_4 \quad \text{and} \quad X_1 < X_5 + X_6 - X_7), \tag{7.83}$$

where $X_1, X_2, \ldots, X_7$ are independent and identically distributed according to the common continuous underlying distribution F in Assumption A3. Since the value of λ_F (7.83) depends on the particular form of the continuous F, we can not use the expression in (7.82) to construct a distribution-free procedure for testing H_0 (7.2). However, Lehmann (1964) showed that $\lambda_F \le \frac{7}{24}$ for all continuous F (see Comment 87). Replacing λ_F in equation (7.82) by this upper bound of $\frac{7}{24}$ yields the expression

$$V_U = \frac{2n - 1 + (k-2)[7(n-2) + 13 - 6n]}{3kn(n-1)}. \tag{7.84}$$

The Doksum test statistic for the conservative test of H_0 (7.2) is then

$$D = \sum_{j=1}^{k} \frac{[H_{j\cdot} - \{(k-1)/2k\}]^2}{(k-1)V_U/2k}. \tag{7.85}$$

For a conservative test (see Comment 85) of

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1 : [\tau_1, \tau_2, \ldots, \tau_k \text{ not all equal}],$$

at the approximate α level of significance,

$$\text{Reject } H_0 \text{ if } D \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise do not reject,} \tag{7.86}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile point of a chi-square distribution with $k-1$ degrees of freedom. To find $\chi^2_{k-1,\alpha}$, we use the R command `qchisq`$(1-\alpha, k-1)$. For example, to find $\chi^2_{3,.05}$, we apply `qchisq(.95, 3)` and obtain $\chi^2_{3,.05} = 7.815$.

Ties

For any Y^i_{uv} (7.77), $1 \leq u < v \leq k$, that is zero, compute T_{uv} and B_{uv} in (7.78) by replacing the associated Ψ^i_{uv} (7.79) with

$$\Psi^{*i}_{uv} = \begin{cases} 1, & \text{if } X_{iu} < X_{iv}, \\ \frac{1}{2}, & \text{if } X_{iu} = X_{iv}, \\ 0, & \text{if } X_{iu} > X_{iv}. \end{cases} \tag{7.87}$$

For ties among $Y^1_{uv}, \ldots, Y^n_{uv}$, use average ranks to compute T_{uv} (7.78).

EXAMPLE 7.11 ***Rounding First Base.***

Consider once again the rounding-first-base data presented in Table 7.1 and discussed in Example 7.1. The reader should already be familiar with the calculations of the paired-data signed rank statistics T_{uv} and sign statistics B_{uv} (see Comment 86) from the materials in Sections 3.1 and 3.4, respectively. We include a detailed calculation of T_{12} and B_{12} in Table 7.25 to illustrate the method for handling zero differences and ties (see Ties and Comment 88).

The statistics B_{12} and T_{12} are obtained by summing the entries in the next-to-last and last columns, respectively, of Table 7.25. We obtain B_{13}, B_{23}, T_{13}, and T_{23} in a similar manner, and the results are

$$B_{13} = 5, \quad B_{23} = 5, \quad T_{13} = 54, \quad \text{and} \quad T_{23} = 30.5.$$

Table 7.25 Calculation of T_{12} and B_{12} for Data in Table 7.1

j	$X_{j1} - X_{j2}$	Y_{12}^j	R_{12}^j	Ψ_{12}^{*j}	$R_{12}^j \Psi_{12}^{*j}$
1	−.10	.10	17	1	17
2	.15	.15	20	0	0
3	−.40	.40	22	1	22
4	.05	.05	9.5	0	0
5	.05	.05	9.5	0	0
6	−.10	.10	17	1	17
7	.00	.00	2.5	$\frac{1}{2}$	1.25
8	−.05	.05	9.5	1	9.5
9	.10	.10	17	0	0
10	.05	.05	9.5	0	0
11	.05	.05	9.5	0	0
12	.10	.10	17	0	0
13	.25	.25	21	0	0
14	.05	.05	9.5	0	0
15	.00	.00	2.5	$\frac{1}{2}$	1.25
16	−.10	.10	17	1	17
17	.00	.00	2.5	$\frac{1}{2}$	1.25
18	−.05	.05	9.5	1	9.5
19	.05	.05	9.5	0	0
20	.05	.05	9.5	0	0
21	.05	.05	9.5	0	0
22	.00	.00	2.5	$\frac{1}{2}$	1.25
				$B_{12} = 8$	$T_{12} = 97$

It then follows from (7.80) that

$$H_{12} = .385, \quad H_{13} = .212, \quad \text{and} \quad H_{23} = .110.$$

From (7.81) and the fact that $H_{uv} = 1 - H_{vu}$, we have

$$\begin{aligned} H_{1\cdot} &= \frac{H_{11} + H_{12} + H_{13}}{3} \\ &= \frac{0 + .385 + .212}{3} = .199, \\ H_{2\cdot} &= \frac{H_{21} + H_{22} + H_{23}}{3} \\ &= \frac{.615 + 0 + .110}{3} = .242, \\ H_{3\cdot} &= \frac{H_{31} + H_{32} + H_{33}}{3} \\ &= \frac{.788 + .890 + 0}{3} = .559. \end{aligned}$$

We next find V_U (7.84) to be

$$V_U = \frac{2(22) - 1 + [7(20) + 13 - 6(22)]}{3(3)(22)(21)} = .015.$$

Substituting these values for $H_{1.}$, $H_{2.}$, $H_{3.}$, and V_U into (7.43) yields

$$D = \frac{\left[.199 - \left(\frac{1}{3}\right)\right]^2 + \left[.242 - \left(\frac{1}{3}\right)\right]^2 + \left[.559 - \left(\frac{1}{3}\right)\right]^2}{2(.015)/6} = 15.5.$$

Referring this value of D to the chi-square distribution with $k - 1 = 2$ degrees of freedom, we use the R command `pchisq(15.5,2)` to find that the lowest significance level at which we would reject H_0 is $1 - .99957 = .00043$ (cf. Example 7.1).

Comments

84. *Motivation for the Test.* Under H_0 (7.2), the $H_{j.}$'s (7.81) tend to be near $(k - 1)/2k$, their common null expectation, and thus the numerator of D (7.85) tends to be small. When the τ's are not all equal, we expect the $H_{j.}$'s to be more disparate, and thus (at least some of) the $[H_{j.} - \{(k-1)/2k\}]^2$ terms tend to be large, yielding a large value of D. This provides partial motivation for procedure (7.86).

85. *Conservative Nature of the Test.* The test defined by (7.86) is neither distribution-free nor asymptotically $(n \to \infty)$ distribution-free. Rather, it is conservative in the sense that (asymptotically) the actual probability of rejecting H_0 (7.2) when it is true tends to be slightly smaller than the nominal level α. This is a consequence of using an upper bound for the parameter λ_F (7.83). See also Comments 87 and 89.

86. *Pairwise Signed Rank and Sign Statistics.* For a given pair of treatments $(u, v), 1 \le u < v \le k$, the statistics T_{uv} and B_{uv} (7.78) are simply the Wilcoxon signed rank and sign statistics, respectively, as discussed in Sections 3.1 and 3.4, respectively, applied to the paired data in treatments u and v. With this relationship in mind, we note that the difference $T_{uv} - B_{uv}$ in the numerator of H_{uv} (7.80) may equivalently be calculated as the number of Walsh averages $(X_{su} - X_{sv} + X_{tu} - X_{tv})/2$, with $1 \le s < t \le n$, that are negative. (See Comment 3.17.)

87. *Bounds for the Parameter* λ_F. The null correlation between two overlapping statistics H_{uv} and H_{uw} defined by (7.80), with $u \ne v, u \ne w$, and $v \ne w$, depends on the parameter λ_F (7.83). This, combined with the fact that λ_F varies with F (Lehmann, 1964), prevents the development of a distribution-free test procedure based on the numerator of D (7.85). Lehmann (1964) showed that $\lambda_F \le \frac{7}{24} (\approx .2917)$ for all continuous F. Replacement of λ_F in expression (7.82) for the null variance of $H_{u.} - H_{v.}$ by the upper bound $\frac{7}{24}$ enables the development of the conservative procedure based on D (7.86). Spurrier (1991) established the lower bound $\lambda_F \ge \frac{89}{315} (\approx .2825)$ for all continuous F. Since the value of λ_F is so narrowly confined between .2825 and .2917 for all continuous F, replacing λ_F by its upper bound or $\frac{7}{24}$ in expression

(7.82) sacrifices little to permit the construction of the conservative test procedure (7.86).

88. *Ties.* The reader may have noted that the method we advocate in Ties for dealing with zero differences, when computing the $T_{uv}(B_{uv})$ signed rank (sign) statistics for use in procedure (7.86), differs from the corresponding directions given for the signed rank (sign) statistic in Section 3.1 (Section 3.4). In Chapter 3, we recommended reducing the sample size by the number of zero differences. This change is initiated in the calculation of D (7.85) in order to keep all of the T_{uv}'s and B_{uv}'s based on the same sample size (n).

89. *Asymptotically Distribution-Free Competitor.* As an alternative to the conservative test procedure (7.86) based on the replacement of the unknown parameter λ_F (7.83) by its upper bound $\frac{7}{24}$, we could instead choose to estimate the value of λ_F from the sample data. Use of a consistent estimator of λ_F in this manner leads to an asymptotically ($n \to \infty$) distribution-free procedure for testing H_0 (7.2), rather than the conservative procedure in (7.86). Lehmann (1964) proposed the estimator $\widehat{\lambda}_F$ of λ_F, where $\widehat{\lambda}_F$ is the proportion of sample sextuples $(\alpha, \beta, \gamma; u, v, w)$ for which the simultaneous inequalities

$$(X_{\alpha u} < X_{\beta u} + X_{\alpha v} - X_{\beta v} \text{ and } X_{\alpha u} < X_{\gamma u} + X_{\alpha w} - X_{\gamma w})$$

are satisfied. In practice, when estimating λ_F, it would normally suffice to check only a subset of the total number of such sample sextuples. Due to the closeness of the upper bound $\frac{7}{24}$ to all values of λ_F, procedure (7.86) is, for all practical purposes, virtually equivalent to Doksum's (1967) asymptotically distribution-free procedure based on estimating λ_F.

Properties

1. *Consistency.* See Doksum (1967) and Hollander and Wolfe (1973, p. 166).
2. *Asymptotic Chi-Square Distribution.* See Doksum (1967).
3. *Efficiency.* See Doksum (1967) and Section 7.16.

Problems

109. Apply procedure (7.86) to the adaptation score data of Table 7.10 (Example 7.4).

110. The Doksum test procedure (7.86) uses between-block information, whereas Friedman's test procedure (7.6) uses only within-block information. Explain.

111. Apply procedure (7.86) to the serumCPK activity data in Table 7.3, Problem 5.

112. Apply procedure (7.86) to the percentage correctly identified consonants data in Table 7.4 (Problem 12).

113. Both the Doksum (7.86) and the Friedman (7.6) procedures are appropriate for testing against general alternatives when we have data from a randomized complete block design with one observation per treatment–block combination. Discuss the relative advantages and disadvantages of the two competing procedures.

7.12 A TEST BASED ON WILCOXON SIGNED RANKS FOR ORDERED ALTERNATIVES IN A RANDOMIZED COMPLETE BLOCK DESIGN (HOLLANDER)

In this section we present a conservative procedure based on pairwise signed ranks for testing H_0 (7.2) against the a priori ordered alternatives H_2 (7.9), corresponding to $\tau_1 \le \tau_2 \le \cdots \le \tau_k$, with at least one strict inequality.

Procedure

For each of the $k(k-1)/2$ pairs of treatments (u, v), with $1 \le u < v \le k$, we compute the signed rank statistic T_{uv}, as defined in (7.78). To compute the Hollander statistic Q, set

$$Y = \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} T_{uv}. \tag{7.88}$$

The null expected value of Y is given by

$$E_0(Y) = \frac{nk(k-1)(n+1)}{8}, \tag{7.89}$$

but the null variance of Y is unknown (see Comment 93) and depends on the particular form of the underlying continuous distribution F in Assumption 3. Thus, a test of H_0 (7.2) based on Y will not be distribution-free. However, a conservative procedure can be developed by using an upper bound for this unknown null variance of Y. Using the R command `CorrUpperBound(n)`, we obtain the value of the upper bound ρ_U^n for the null correlation between two overlapping signed rank statistics based on n observations. An upper bound for the null variance of Y (7.88) is then given by

$$\text{var}_U(Y) = \frac{nk(n+1)(2n+1)(k-1)\{3+2(k-2)\rho_U^n\}}{144}. \tag{7.90}$$

The Hollander test statistic for the conservative test of H_0 (7.2) is then

$$Q = \frac{Y - E_0(Y)}{\{\text{var}_U(Y)\}^{1/2}}, \tag{7.91}$$

with the expressions for $E_0(Y)$ and $\text{var}_U(Y)$ given in (7.89) and (7.90), respectively. For a conservative test (see Comment 92) of

$$H_0 : [\tau_1 = \cdots = \tau_k]$$

versus the ordered alternatives

$$H_2 : [\tau_1 \le \tau_2 \le \cdots \le \tau_k, \text{ with at least one strict inequality}],$$

at the approximate α level of significance,

$$\text{Reject } H_0 \text{ if } Q \ge z_\alpha; \quad \text{otherwise do not reject.} \tag{7.92}$$

Ties

See Ties of Section 7.11 and Comment 88.

EXAMPLE 7.12 ***Effect of Weight on Forearm Tremor Frequency.***

The data in Table 7.26 are based on a subset of the data obtained by Fox and Randall (1970) in their study of forearm tremor. Each entry in the table is the mean of five experimental values of tremor frequency. We identify treatment 1 with 7.5 lb, treatment 2 with 5 lb, treatment 3 with 2.5 lb, treatment 4 with 1.25 lb, and treatment 5 with 0 lb, and use procedure (7.92) to test H_0 (7.2) versus the ordered alternatives H_2 (7.9), which specify that adding mass decreases the tremor frequency.

Calculations similar to those presented in Example 7.11 yield

$$\begin{aligned} &T_{12} = 18.5, \quad T_{13} = 21, \quad T_{14} = 21, \quad T_{15} = 21, \quad T_{23} = 20, \\ &T_{24} = 21, \quad T_{25} = 21, \quad T_{34} = 21, \quad T_{35} = 21, \quad T_{45} = 21. \end{aligned} \tag{7.93}$$

From (7.88), we obtain

$$Y = T_{12} + T_{13} + T_{14} + T_{15} + T_{23} + T_{24} + T_{25} + T_{34} + T_{35} + T_{45} = 206.5.$$

From the R command `CorrUpperBound(6)`, we find $\rho_U^6 = .452$, and evaluating (7.89) and (7.90) gives

$$E_0(Y) = \frac{5(4)(6)(7)}{8} = 105,$$

$$\mathrm{Var}_U(Y) = \frac{6(7)(13)(5)(4)\{3 + 6(.452)\}}{144} = 433.2.$$

From (7.91), we then have

$$Q = \frac{206.5 - 105}{[433.2]^{1/2}} = 4.88.$$

Table 7.26 Forearm Tremor Frequency (Hz) as a Function of Weight Applied at the Wrist

Treatment	1	2	3	4	5
			Weight (lb)		
Subject	7.5	5	2.5	1.25	0
1	2.58	2.63	2.62	2.85	3.01
2	2.70	2.83	3.15	3.43	3.47
3	2.78	2.71	3.02	3.14	3.35
4	2.36	2.49	2.58	2.86	3.10
5	2.67	2.96	3.08	3.32	3.41
6	2.43	2.50	2.85	3.06	3.07

Source: J. R. Fox and J. E. Randall (1970).

Using the R command `pnorm(·)`, we see that the lowest approximate level at which we would reject H_0 with these data is $P_0(Q \geq 4.88) \approx 1 - \texttt{pnorm(4.88)} = 1 - .99999947 = .00000053$. Thus there is very strong evidence (over the range of weights considered in the study) that the tremor frequency does decrease as the applied weight increases.

Comments

90. *Motivation for the Test.* Note that the statistic Y (7.88) is designed to guard against the postulated ordered alternatives H_2 (7.9). Consider the case $k = 3$. Then $Y = T_{12} + T_{13} + T_{23}$, and if $\tau_1 < \tau_2 < \tau_3$, each of T_{12}, T_{13}, and T_{23} would tend to be larger than $n(n+1)/4$, their common null expectation. Thus, Y would tend to be large, as desired. Contrast this with a situation in which we suspect (and design the test for) the alternative $\tau_1 < \tau_2 < \tau_3$, but in actuality, we have $\tau_3 < \tau_2 < \tau_1$. In this case, each of T_{12}, T_{13}, and T_{23} would tend to be small. This provides partial motivation for procedure (7.92).

91. *Non-Distribution-Free Property of Y (7.88).* Consider the Y (7.88) statistic for testing against ordered alternatives in the two-way layout (7.1) in relation to Jonckheere's J (6.13) statistic for testing against ordered alternatives in the one-way layout (6.1). The statistic J is the sum $\sum_{u<v}^{k} U_{uv}$ of two-sample Mann–Whitney statistics U_{uv} (or, equivalently, Wilcoxon rank sum statistics), where each U_{uv} is distribution-free under H_0 (6.2). The statistic Y is a sum $\sum_{u<v}^{k} T_{uv}$ of the paired-sample Wilcoxon signed rank statistics T_{uv}, where each T_{uv} is distribution-free under H_0 (7.2). Although J itself is also distribution-free under H_0 (6.2), Y is not distribution-free under H_0 (7.2) when $k > 2$. (For $k = 2$, Y reduces to T_{12}, which is distribution-free.) See Hollander (1967a) for details of the non-distribution-free character of Y.

92. *Conservative Nature of the Test.* The test defined in (7.91) is neither distribution-free nor asymptotically ($n \to \infty$) distribution-free. Rather, it is conservative in the sense that (asymptotically) the actual probability of rejecting H_0 (7.2) when it is true tends to be smaller than the nominal level α. This is a direct consequence of using an upper bound $\text{var}_U(Y)$ to replace the unknown null variance of Y. Also see Comment 94.

93. *Asymptotic Null Variance of Y.* Hollander (1967a) showed that the asymptotic ($n \to \infty$) null variance of Y (7.88) has the form

$$\text{var}_0(Y) = \frac{nk(n+1)(2n+1)(k-1)\{3 + 2(k-2)\rho^*\}}{144},$$

where ρ^* is the limiting ($n \to \infty$) null correlation between two overlapping signed rank statistics T_{12} and T_{13}. This limiting correlation can also be expressed as

$$\rho^* = 12\lambda_F - 3, \tag{7.94}$$

where λ_F is defined by (7.83). In forming the Q test statistic (7.91) for the conservative test procedure (7.92), we replace ρ^* by its upper bound ρ_U^n.

94. *Asymptotically Distribution-Free Competitor.* As an alternative to the conservative test procedure (7.92) based on the use of the upper bound ρ_U^n, we could instead replace ρ^* (7.94) by a consistent estimator $\widehat{\rho}$ based on the sample data. Use of a consistent estimator of ρ^* in this manner leads to an asymptotically ($n \to \infty$) distribution-free procedure for testing H_0 (7.2) rather than the conservative procedure in (7.92). Hollander suggested such an approach to this problem based on the consistent estimator $\widehat{\rho} = 12\widehat{\lambda}_F - 3$, where $\widehat{\lambda}_F$ is defined in Comment 89. Due to the closeness of the upper bound $\frac{7}{24}$ to all values of λ_F, procedure (7.92) is, for all practical purposes, virtually equivalent to Hollander's (1967a) asymptotically distribution-free procedure based on estimating λ_F.

Properties

1. *Consistency.* The test defined by (7.92) is consistent against the ordered alternatives (7.9). See Hollander (1967a) and Hollander and Wolfe (1973, p. 170).
2. *Asymptotic Normality.* See Hollander (1967a).
3. *Efficiency.* See Hollander (1967a) and Section 7.16.

Problems

114. Apply the Q (7.92) test to the metronome data of Table 7.6. Use the postulated ordering $\tau_R < \tau_A < \tau_N$.

115. The Hollander test procedure (7.92) uses between-block information, but Page's test procedure (7.11) uses only within-block information. Explain.

116. Apply procedure (7.92) to the shelterbelt data in Table 7.7 (Problem 19).

117. Apply procedure (7.92) to the cotton strength index data in Table 7.5 (Example 7.2). Compare with the result from the use of Page's test in Example 7.2.

118. Both the Hollander (7.92) and the Page (7.11) procedures are appropriate for testing against ordered alternatives when we have data from a randomized complete block design with one observation per treatment–block combination. Discuss the relative advantages and disadvantages of the two competing procedures.

7.13 APPROXIMATE TWO-SIDED ALL-TREATMENTS MULTIPLE COMPARISONS BASED ON SIGNED RANKS (NEMENYI)

In this section we present a multiple comparison procedure based on Wilcoxon signed rank statistics that is designed to make decisions about individual differences between pairs of treatment effects (τ_u, τ_v), for $u < v$, in a setting where general alternatives H_1 (7.3) are of interest. Thus, the multiple comparison procedure of this section would generally be applied to two-way layout data (with one observation per cell) *after* rejection of H_0 (7.2) with the Doksum–Lehmann procedure from Section 7.11. In this setting it is important to reach conclusions about all $k(k-1)/2$ pairs of treatment effects and these conclusions are naturally two sided in nature.

Procedure

For $1 \le u < v \le k$, let T_{uv} be the signed rank statistic (7.78) between treatments u and v. Calculate the $k(k-1)/2$ statistics

$$T'_{uv} = \max\{T_{uv}, [n(n+1)/2] - T_{uv}\}, 1 \le u < v \le k. \tag{7.95}$$

At an approximate (see Comment 95) experimentwise error rate of α, the two-sided signed rank multiple comparison procedure reaches its $k(k-1)/2$ pairwise decisions, corresponding to each (τ_u, τ_v) pair, for $1 \le u < v \le k$, by the criterion

$$\text{Decide } \tau_u \ne \tau_v \text{ if } T'_{uv} \ge t'_\alpha; \quad \text{otherwise decide } \tau_u = \tau_v, \tag{7.96}$$

where the constant t'_α is chosen to make the experimentwise error rate approximately equal to α; that is, t'_α satisfies the restriction

$$P_0\{T'_{uv} < t'_\alpha, u = 1, \ldots, k-1 \quad \text{and} \quad v = u+1, \ldots, k\} \approx 1 - \alpha, \tag{7.97}$$

where the probability $P_0(.)$ is computed under H_0 (7.2). Equation (7.97) stipulates that the $k(k-1)/2$ inequalities $T'_{uv} < t'_\alpha$, corresponding to all pairs (u, v) of treatments with $u < v$, hold simultaneously with approximate probability $1 - \alpha$ when H_0 (7.2) is true. Selected approximate values of t'_α can be found from the relationship

$$t'_\alpha \approx \left[\frac{n(n+1)}{4}\right] + \left[\frac{n(n+1)(2n+1)}{48}\right]^{1/2} q_\alpha, \tag{7.98}$$

where q_α is the upper αth percentile point for the distribution of the range of k independent $N(0, 1)$ variables. To find q_α for k treatments and a specified experimentwise error rate α, we use the R command `cRangeNor(`α`, k)`. For example, to find $q_{.005}$ for $k = 6$ treatments, we apply `cRangeNor(.005, 6)` and obtain $q_{.005} = 5.033$ for $k = 6$.

Ties

See Ties of Section 7.11 and Comment 88.

EXAMPLE 7.13 ***Rounding First Base.***

We illustrate procedure (7.96) using the approximation (7.98) with the rounding-first-base data of Table 7.1. In Example 7.11, we found

$$T_{12} = 97, \quad T_{13} = 54, \quad \text{and} \quad T_{23} = 30.5.$$

From (7.95), we obtain

$$T'_{12} = \max\{97, 253 - 97\} = 156, \quad T'_{13} = \max\{54, 253 - 54\} = 199,$$

$$T'_{23} = \max\{30.5, 253 - 30.5\} = 222.5.$$

With an experimentwise error rate of $\alpha = .01$ and $k = 3$, we use `cRangeNor(.01,3)` to find $q_{.01} = 4.12$ for $k = 3$. Thus, with approximation (7.98), the inequality in (7.96) reduces to

$$T'_{uv} \geq t'_{.01} \approx \left[\frac{22(23)}{4}\right] + \left[\frac{22(23)(45)}{48}\right]^{1/2} (4.12) = 216.2,$$

and procedure (7.96) becomes

$$\text{Decide } \tau_u \neq \tau_v \text{ if } T'_{uv} \geq 216.2, \quad 1 \leq u < v \leq 3.$$

Since $T'_{12} < 216.2$, $T'_{13} < 216.2$, and $T'_{23} \geq 216.2$, only the narrow angle (treatment 2) and wide angle (treatment 3) running methods differ significantly at the approximate .01 experimentwise error rate using the signed rank procedure (7.96).

At this point, the reader may have noticed that, at the approximate .01 experimentwise error rate, the signed rank analysis in this example yields a conclusion different from the corresponding analysis based on the Friedman rank sums performed in Example 7.3. Since the analyses are based on different rankings and different statistics, the reader should not be shocked. It is instructive to note that if, for example, the multiple comparisons were made at an approximate .10 experimentwise error rate, the two procedures would agree in the sense that differences between treatments 2 and 3 and between treatments 1 and 3 would be declared significant under both analyses.

Comments

95. *Non-Distribution-Free Property.* Procedure (7.96), using approximation (7.98), is neither distribution-free nor asymptotically distribution-free. Nemenyi (1963) proposed this procedure under the assumptions that (a) the statistic $\max\{T'_{uv}, 1 \leq u < v \leq k\}$ is distribution-free and (b) the limiting ($n \to \infty$) null correlation between T_{12} and T_{13} (say) is close to $\frac{1}{2}$. Assumption (a) is incorrect, but the reasonableness of assumption (b) is supported by the values of λ_F, for various distributions F, obtained by Lehmann (1964), Hollander (1966), and Obenchain (1969). (See also Comments 87 and 93.)
96. *Independence from Observations for Other Noninvolved Treatments.* The value of T'_{uv}, the statistic used in the decision relating to τ_u and τ_v, does not depend on the observation values from the other $k - 2$ treatments. Thus, the signed ranks procedure (7.96) eliminates a difficulty encountered with the corresponding multiple comparison procedures (7.25) and (7.27) of Section 7.3 based on the Friedman rank sums. (See Comment 30.)

Properties

1. *Efficiency.* See Section 7.16.

Problems

119. Apply procedure (7.96) to the serum CPK activity data in Table 7.3 (Problem 5).

120. Apply procedure (7.96) to the Hebb–Williams EPT data in Table 7.9 (Problem 28).

121. Both procedures (7.27) and (7.96) are appropriate multiple comparison procedures when we have data from a randomized complete block design with one observation per treatment–block combination, and we are interested in two-sided comparisons between all treatments. Discuss the relative advantages and disadvantages of the two competing procedures.

122. Apply procedure (7.96) to the percentage correctly identified consonants data in Table 7.4 (Problem 12).

7.14 APPROXIMATE ONE-SIDED TREATMENTS-VERSUS-CONTROL MULTIPLE COMPARISONS BASED ON SIGNED RANKS (HOLLANDER)

In this section we turn our attention to a multiple comparison procedure based on the Wilcoxon signed rank statistics that is designed to make decisions about individual differences between the median effect for a single, baseline control population, and the median effects of each of the remaining $k-1$ treatments. This treatments- versus-control multiple comparison procedure can be applied to two-way layout data (with one observation per cell) *after* rejection of H_0 (7.2) with either the Doksum–Lehmann or the Hollander procedure discussed in Sections 7.11 and 7.12, respectively. Its application leads to conclusions about the differences between each of the $k-1$ treatment effects and the control effect and these conclusions are naturally one sided in nature.

Procedure

For simplicity of notation, we let treatment 1 assume the role of the single, baseline control. For each of the $k-1$ treatments $u = 2, \ldots, k$, we compute the signed rank statistic T_{1u} (7.78) between the control treatment 1 and treatment u. At an approximate (see Comment 98) experimentwise error rate of α, the one-sided treatments-versus-control signed rank multiple comparison procedure reaches its $k-1$ pairwise decisions, corresponding to each (τ_1, τ_u) pair, for $u = 2, \ldots, k$, by the criterion

$$\text{Decide } \tau_u > \tau_1 \text{ if } T_{1u} \geq t_\alpha^*; \quad \text{otherwise decide } \tau_u = \tau_1, \tag{7.99}$$

where the constant t_α^* is chosen to make the experimentwise error rate approximately equal to α; that is, t_α^* satisfies the restriction

$$P_0\{T_{1u} < t_\alpha^*, u = 2, \ldots, k\} \approx 1 - \alpha, \tag{7.100}$$

where the probability $P_0(.)$ is computed under H_0 (7.2). Equation (7.100) stipulates that the $k-1$ inequalities $T_{1u} < t_\alpha^*$, corresponding to each treatment paired with the control, hold simultaneously with approximate probability $1-\alpha$ when H_0 (7.2) is true. Selected approximate values of t_α^* can be found from the relationship

$$t_\alpha^* \approx \left[\frac{n(n+1)}{4}\right] + \left[\frac{n(n+1)(2n+1)}{24}\right]^{1/2} m_{\alpha,\rho^*}^*, \tag{7.101}$$

where m_{α,ρ^*}^* is the upper αth percentile point for the distribution of the maximum of $(k-1)N(0,1)$ variables with common correlation ρ^* equal to the upper bound ρ_U^n for

the null correlation between two overlapping signed rank statistics based on n observations. The upper bound ρ_U^n for signed rank statistics based on n observations is found from the R command `CorrUpperBound(n)`. To find m^*_{α,ρ^*} for k treatments and a specified experimentwise error rate α, we use the R command `cMaxCorrNor(`α`,k,rho.hat)`. For example, to find $m^*_{.02337,.3}$ for $k = 5$ treatments and correlation $\rho^* = .3$, we apply `cMaxCorrNor(.02337,5,.3)` and obtain $m^*_{.02337,.3} = 2.50$. (For a discussion of how to adjust procedure (7.99) for settings where it is of interest to decide whether the treatment effects are *smaller* than the control effect, see Comment 97.)

Ties

See Ties of Section 7.11 and Comment 88.

EXAMPLE 7.14 ***Effect of Weight on Forearm Tremor Frequency.***

We use the tremor data of Table 7.26 to illustrate procedure (7.99) using the approximation (7.101). We relabel the treatments so that the no-weight (0 lb) treatment assumes the role of the control. To make this clear in the ensuing computations, we reproduce Table 7.26 as Table 7.26′ with the new treatment designations.

We illustrate the one-sided decisions of $\tau_u = \tau_1$ versus $\tau_u < \tau_1$, $u = 2, \ldots, 5$. We see from Comment 97 that our procedure is based on (7.99) with $T_{u1} = [n(n+1)/2] - T_{1u}$ replacing T_{1u} in the left-hand side of the inequality in (7.99). From the relabeling in Table 7.26′ and the basic computations in Example 7.12, we obtain

$$T_{12} = 0, \quad T_{13} = 0, \quad T_{14} = 0, \quad \text{and} \quad T_{15} = 0,$$

which, in turn, implies

$$T_{21} = 21 - T_{12} = 21, \quad T_{31} = 21 - T_{13} = 21,$$

$$T_{41} = 21 - T_{14} = 21, \quad T_{51} = 21 - T_{15} = 21.$$

From the R command `CorrUpperBound(6)`, we find $\rho_U^6 = .452$. With an approximate experimentwise error rate of $\alpha = .10$, we then use the R command `cMaxCorrNor(.10,`

Table 7.26′ Forearm Tremor Frequency (Hz) as a Function of Weight Applied at the Wrist

Treatment	1	2	3	4	5
			Weight (lb)		
Subject	0	1.25	2.5	5	7.5
1	3.01	2.85	2.62	2.63	2.58
2	3.47	3.43	3.15	2.83	2.70
3	3.35	3.14	3.02	2.71	2.78
4	3.10	2.86	2.58	2.49	2.36
5	3.41	3.32	3.08	2.96	2.67
6	3.07	3.06	2.85	2.50	2.43

Source: J. R. Fox and J. E. Randall (1970).

5, .452) with $\rho^* = .452$ to find $m^*_{.10,.452} = 1.935$. Thus, with approximation (7.101), the inequality in (7.99) for our one-sided decisions reduces to

$$T_{u1} \geq t^*_{.10} \approx \left[\frac{6(7)}{4}\right] + \left[\frac{6(7)(13)}{24}\right]^{1/2}(1.935) = 19.73,$$

and procedure (7.99) for these one-sided decisions becomes

$$\text{Decide } \tau_u < \tau_1 \text{ if } T_{u1} \geq 19.73, \quad u = 2, \ldots, 5.$$

Since $T_{21} = T_{31} = T_{41} = T_{51} = 21 > 19.73$, the signed rank procedure (7.99) with an approximate .10 experimentwise error rate concludes that all four weight levels (treatments) yield significantly smaller forearm tremor frequencies than does the zero weight control.

Comments

97. *Opposite Direction Decisions.* Procedure (7.99) is designed for the one-sided situation in which the relevant decisions are $\tau_u = \tau_1$ versus $\tau_u > \tau_1, u = 2, \ldots, k$. To treat the analogous one-sided case of $\tau_u = \tau_1$ versus $\tau_u < \tau_1$, $u = 2, \ldots, k$, we simply replace T_{1u} by $T_{u1} = [n(n+1)/2] - T_{1u}$, in the left-hand side of the inequality in (7.99).
98. *Non-Distribution-Free Property*. Procedure (7.99), using approximation (7.101), is neither distribution-free nor asymptotically distribution-free. However, it is generally conservative in that the attained approximate experimentwise error rate associated with procedure (7.99) tends to be slightly lower than the nominally stipulated rate. Due to the closeness of the upper bound $\frac{7}{24}$ to all values of λ_F, procedure (7.99) is, for all practical purposes, virtually equivalent to Hollander's (1966) asymptotically distribution-free procedure based on estimating λ_F.
99. *Simplification of Approximation.* One of the disadvantages of the approximation to t^*_α provided in (7.101) is that it requires obtaining the value of m^*_{α,ρ^*} for common correlation $\rho^* = \rho^n_U$. To simplify matters, one could use the further approximation associated with replacing ρ^n_U by its asymptotic ($n \to \infty$) limit of $\frac{1}{2}$. The approximation in (7.101) would then use the proper value of $m^*_{\alpha,\frac{1}{2}}$ for common correlation $\rho^* = \frac{1}{2}$.
100. *Asymptotically Distribution-Free Competitor.* As an alternative to the conservative one-sided multiple comparison procedure (7.99) based on the use of the upper bound ρ^n_U in approximation (7.101), we could instead use a consistent estimator $\hat{\rho}$ of the null correlation between two overlapping signed rank statistics based on n observations. The value of m^*_{α,ρ^*} used in approximation (7.101) would then correspond to this estimate ($\hat{\rho}$) of the null correlation rather than the upper bound ρ^n_U. Use of a consistent estimator $\hat{\rho}$ in this manner leads to an asymptotically ($n \to \infty$) distribution-free one-sided multiple comparison procedure, rather than the conservative procedure in (7.99). Hollander (1966) suggested such an approach based on the consistent estimator $\hat{\rho} = 12\hat{\lambda}_F - 3$, where $\hat{\lambda}_F$ is defined in Comment 89.

101. *Two-Sided Treatments-versus-Control Multiple Comparison Procedure*. The multiple comparison procedure (7.99) of this section is one sided by nature, resulting in decisions between $\tau_u = \tau_1$ and $\tau_u > \tau_1$ for every $u = 2, \ldots, k$ (or between $\tau_u = \tau_1$ and $\tau_u < \tau_1$ for every $u = 2, \ldots, k$, as noted in Comment 97). We view such one-sided comparisons to be the most natural approach for treatments-versus-control settings. In such situations, we are generally interested in seeing which, if any, of the proposed new treatments are better than a standard control or placebo. In most practical applications, *better* is synonymous with one-sided comparisons (all in one direction or all in the other) and thus our emphasis on such procedures in this section. However, a two-sided treatments-versus-control analog to procedure (7.99) has been developed in the literature and corresponds to the criterion

$$\text{Decide } \tau_u \neq \tau_1 \text{ if } T'_{1u} \geq t^{**}_{\alpha}; \quad \text{otherwise decide } \tau_u = \tau_1, \tag{7.102}$$

where the T'_{1u}'s are defined by (7.95) and the constant t^{**}_{α} is chosen to make the experimentwise error rate approximately equal to α; that is,

$$P_0\{T'_{1u} < t^{**}_{\alpha}, u = 2, \ldots, k\} \approx 1 - \alpha,$$

where the probability $P_0(.)$ is computed under H_0 (7.2). One approximation for t^{**}_{α} sets

$$t^{**}_{\alpha} \approx \left[\frac{n(n+1)}{4}\right] + \left[\frac{n(n+1)(2n+1)}{24}\right]^{1/2} v^*_{\alpha}, \tag{7.103}$$

where v^*_{α} is the upper αth percentile of the maximum absolute value of $(k-1)N(0,1)$ random variables with common correlation $\frac{1}{2}$. Selected values of v^*_{α} can be obtained from Dunnett (1964).

102. *Independence from Observations for Other Noninvolved Treatments.* The value of T_{1u}, the statistic used in the decision relating to τ_u and τ_1, does not depend on the observation values from the other $k-2$ treatments. Thus, the signed ranks procedure (7.99) eliminates a difficulty encountered with the corresponding one-sided multiple comparison procedures (7.28) and (7.30) of Section 7.4 based on the Friedman rank sums. (See Comment 38.)

Properties

1. *Efficiency.* See Section 7.16.

Problems

123. Apply an appropriate one-sided signed rank multiple comparison procedure (see (7.99) and Comment 97) to the rhythmicity data of Table 7.6 in Problem 15, letting the condition *N* serve as the control.

124. Consider the serum CPK activity data from Problem 5. Treating preexercise as a control and ignoring the peak psychotic period data, apply procedure (7.99) to decide if there is statistical evidence of increased serum CPK activity either 19 or 42 h after exercise.

125. Both procedures (7.30) and (7.99) are appropriate multiple comparison procedures when we have data from a randomized complete block design with one observation per treatment–block combination and we are interested in one-sided comparisons between $(k-1)$ treatments and a single, baseline control. Discuss the relative advantages and disadvantages of the two competing procedures.

126. Treating condition A as a control, apply procedure (7.99) to the percentage correctly identified consonants data in Table 7.4 (Problem 12).

7.15 CONTRAST ESTIMATION BASED ON THE ONE-SAMPLE HODGES–LEHMANN ESTIMATORS (LEHMANN)

In this section we describe how to use the Hodges–Lehmann estimators based on the appropriate Walsh averages to construct estimators of a contrast θ (7.32 and 7.34) in the treatment effects $\tau_1, \ldots, \tau_k$. For a given setting, decisions about which contrasts to estimate can be related either to a priori interest in particular linear combinations of the τ's or to the result of one of the multiple comparison procedures discussed in Sections 7.3, 7.4, 7.13, and 7.14.

Procedure

Let θ be an arbitrary contrast (7.32 and 7.34) in the treatment effects $\tau_1, \ldots, \tau_k$. For each pair of treatments $(u, v), u \neq v = 1, \ldots, k$, compute the differences D_{uv}^i (7.36), $i = 1, \ldots, n$, between the treatment u and treatment v observations for each of the n blocks. For each (u, v) pair, obtain the values of the $n(n+1)/2$ Walsh averages for these sample differences, namely,

$$\frac{D_{uv}^i + D_{uv}^j}{2}, \quad 1 \leq i \leq j \leq n. \tag{7.104}$$

Let W_{uv} be the median of the Walsh averages associated with the $u - v$ treatment differences; that is,

$$W_{uv} = \text{median} \left\{ \frac{D_{uv}^i + D_{uv}^j}{2}, 1 \leq i \leq j \leq n \right\}, \quad u \neq v = 1, \ldots, k. \tag{7.105}$$

(Since $W_{vu} = -W_{uv}$, we need to calculate only the $k(k-1)/2$ values W_{uv} corresponding to $u < v$.) Note that each W_{uv} is a Hodges–Lehmann estimator of the form considered in Section 3.2, applied here to the $X_{iu} - X_{iv}$ differences. For example, W_{23} is the median of the $n(n+1)/2$ Walsh averages of the form $[D_{23}^i + D_{23}^j]/2$, $1 \leq i \leq j \leq n$, and can be viewed as an "unadjusted" estimator (see Comments 7.103 and 7.104) of the simple contrast $\tau_2 - \tau_3$.

Next, we compute

$$W_{u\cdot} = \sum_{j=1}^{k} \frac{W_{uj}}{k}, \quad u = 1, \ldots, k, \tag{7.106}$$

where we note that $W_{uu} = 0$ for $u = 1, \ldots, k$. Setting

$$\hat{\Delta}_{uv} = W_{u.} - W_{v.}, \tag{7.107}$$

the adjusted estimator of θ is given by

$$\hat{\theta} = \sum_{j=1}^{k} a_j W_{j.}, \tag{7.108}$$

or, equivalently,

$$\hat{\theta} = \sum_{h=1}^{k} \sum_{j=1}^{k} d_{hj} \hat{\Delta}_{hj}. \tag{7.109}$$

(See (7.35) for the relationship between the d's and the a's.)

EXAMPLE 7.15 ***Rounding First Base.***

In Example 7.5, we obtained the Doksum estimator of the contrast $\theta = \tau_{\text{roundout}} - \tau_{\text{wide angle}} = \tau_1 - \tau_3$ relating to the rounding-first-base data of Table 7.1. We now use (7.108) to obtain the Lehmann estimator of the same contrast. To evaluate W_{12}, defined by (7.105), note that $W_{12} = \text{median}\{[D_{12}^i + D_{12}^j]/2, 1 \le i \le j \le 22\}$, where D_{12}^i and D_{12}^j are defined by (7.36). The $D_{12}^1, \ldots, D_{12}^{22}$ values are exhibited in Table 7.11. Letting $F_{12}^{(1)} \le \cdots \le F_{12}^{(253)}$ denote the 253 ordered $[D_{12}^i + D_{12}^j]/2$ values, we find

$$F_{12}^{(1)} = -.4, \quad F_{12}^{(2)} = F_{12}^{(3)} = F_{12}^{(4)} = -.25 \quad F_{12}^{(5)} = F_{12}^{(6)} = -.225,$$

$$F_{12}^{(7)} = \cdots = F_{12}^{(10)} = -.2, \qquad F_{12}^{(11)} = \cdots = F_{12}^{(18)} = -.175,$$

$$F_{12}^{(19)} = F_{12}^{(20)} = -.15, \qquad F_{12}^{(21)} = -.125, F_{12}^{(22)} = \cdots = F_{12}^{(27)} = -.1,$$

$$F_{12}^{(28)} = \cdots = F_{12}^{(34)} = -.075, \qquad F_{12}^{(35)} = \cdots = F_{12}^{(49)} = -.05,$$

$$F_{12}^{(50)} = \cdots = F_{12}^{(81)} = -.025, \qquad F_{12}^{(82)} = \cdots = F_{12}^{(113)} = 0,$$

$$F_{12}^{(114)} = \cdots = F_{12}^{(152)} = .025, \qquad F_{12}^{(153)} = \cdots = F_{12}^{(198)} = .05,$$

$$F_{12}^{(199)} = \cdots = F_{12}^{(221)} = .075, \qquad F_{12}^{(222)} = \cdots = F_{12}^{(234)} = .1,$$

$$F_{12}^{(235)} = \cdots = F_{12}^{(240)} = .125, \qquad F_{12}^{(241)} = \cdots = F_{12}^{(249)} = .15,$$

$$F_{12}^{(250)} = F_{12}^{(251)} = .175, \qquad F_{12}^{(252)} = .2, F_{12}^{(253)} = .25.$$

Thus,

$$W_{12} = F_{12}^{(127)} = .025.$$

To evaluate W_{13}, we use the equation $W_{13} = \text{median}\ \{[D_{13}^i + D_{13}^j]/2, 1 \le i \le j \le 22\}$, where D_{13}^i and D_{13}^j are defined by (7.36). The $D_{13}^1, \ldots, D_{13}^{22}$ values are exhibited in

Table 7.11. Letting $F_{13}^{(1)} \le \cdots \le F_{13}^{(253)}$ denote the 253 ordered $[D_{13}^i + D_{13}^j]/2$ values, we have

$$F_{13}^{(1)} = -.3, \quad F_{13}^{(2)} = F_{13}^{(3)} = -.225, \qquad F_{13}^{(4)} = -.2, \quad F_{13}^{(5)} = -.175,$$
$$F_{13}^{(6)} = F_{13}^{(7)} = F_{13}^{(8)} = -.15, \qquad F_{13}^{(9)} = \cdots = F_{13}^{(12)} = -.125,$$
$$F_{13}^{(13)} = \cdots = F_{13}^{(19)} = -.1, \qquad F_{13}^{(20)} = \cdots = F_{13}^{(25)} = -.075,$$
$$F_{13}^{(26)} = \cdots = F_{13}^{(33)} = -.05, \qquad F_{13}^{(34)} = \cdots = F_{13}^{(46)} = -.025,$$
$$F_{13}^{(47)} = \cdots = F_{13}^{(62)} = 0, \qquad F_{13}^{(63)} = \cdots = F_{13}^{(77)} = .025,$$
$$F_{13}^{(78)} = \cdots = F_{13}^{(94)} = .05, \qquad F_{13}^{(95)} = \cdots = F_{13}^{(108)} = .075,$$
$$F_{13}^{(109)} = \cdots = F_{13}^{(131)} = .1, \qquad F_{13}^{(132)} = \cdots = F_{13}^{(157)} = .125,$$
$$F_{13}^{(158)} = \cdots = F_{13}^{(190)} = .15, \qquad F_{13}^{(191)} = \cdots = F_{13}^{(217)} = .175,$$
$$F_{13}^{(218)} = \cdots = F_{13}^{(238)} = .2, \qquad F_{13}^{(239)} = \cdots = F_{13}^{(247)} = .225,$$
$$F_{13}^{(248)} = \cdots = F_{13}^{(253)} = .25.$$

Thus,

$$W_{13} = F_{13}^{(127)} = .1.$$

In the same way, we calculate W_{23} by using the $D_{23}^1, \ldots, D_{23}^{22}$ values in Table 7.11 and the fact that $W_{23} = \text{median } \{[D_{23}^i + D_{23}^j]/2, 1 \le i \le j \le 22\}$. Letting $F_{23}^{(1)} \le \cdots \le F_{23}^{(253)}$ denote the 253 ordered $[D_{23}^i + D_{23}^j]/2$ values, we see that

$$F_{23}^{(1)} = -.1, F_{23}^{(2)} = \cdots = F_{23}^{(5)} = -.075, \qquad F_{23}^{(6)} = \cdots = F_{23}^{(15)} = -.05,$$
$$F_{23}^{(16)} = \cdots = F_{23}^{(19)} = -.025, \qquad F_{23}^{(20)} = \cdots = F_{23}^{(42)} = 0,$$
$$F_{23}^{(43)} = \cdots = F_{23}^{(73)} = .025, \qquad F_{23}^{(74)} = \cdots = F_{23}^{(98)} = .05,$$
$$F_{23}^{(99)} = \cdots = F_{23}^{(138)} = .075, \qquad F_{23}^{(139)} = \cdots = F_{23}^{(178)} = .1,$$
$$F_{23}^{(179)} = \cdots = F_{23}^{(211)} = .125, \qquad F_{23}^{(212)} = \cdots = F_{23}^{(238)} = .15,$$
$$F_{23}^{(239)} = \cdots = F_{23}^{(247)} = .175, \qquad F_{23}^{(248)} = \cdots = F_{23}^{(253)} = .2.$$

Thus,

$$W_{23} = F_{23}^{(127)} = .075.$$

From (7.106), we find

$$W_{1.} = \frac{W_{11} + W_{12} + W_{13}}{3}$$
$$= \frac{0 + .025 + .1}{3} = .0417,$$

$$W_{2\cdot} = \frac{W_{21} + W_{22} + W_{23}}{2}$$
$$= \frac{-.025 + 0 + .075}{3} = 0.167,$$

and

$$W_{3\cdot} = \frac{W_{31} + W_{32} + W_{33}}{3} = \frac{-.1 - .075 + 0}{3} = -.0583.$$

Note that in calculating $W_{2\cdot}$ and $W_{3\cdot}$, we use the relationship $W_{uv} = -W_{vu}$.

The Lehmann estimator $\hat{\theta}$ is now obtained from (7.108) by noting that $a_1 = 1$, $a_2 = 0$, and $a_3 = -1$, so that

$$\hat{\theta} = W_{1\cdot} - W_{3\cdot} = .0417 - (-.0583) = .10.$$

For these data, the adjusted estimator $W_{1\cdot} - W_{3\cdot}$ agrees with the unadjusted estimator W_{13}. However, we do note that the value of the Lehmann estimator $\hat{\theta} = .10$ differs from that of the Doksum estimator $\tilde{\theta} = .133$ (see Example 7.5) for these rounding-first-base data.

Comments

103. *Unadjusted Estimator*. The unadjusted estimator W_{uv} (7.105) of $\Delta_{uv} = \tau_u - \tau_v$ is simply the estimator associated with the signed rank test and discussed in Section 3.2.

104. *Ambiguities with the Unadjusted Estimators.* The unadjusted estimators W_{uv} (7.105) are incompatible, leading to possible ambiguities in contrast estimation because they do not satisfy the linear relations that are satisfied by the contrasts they estimate. We have encountered this difficulty before (see Comments 6.77 and 7.42). The adjusted estimators $\hat{\Delta}_{uv}$ (7.107) are compatible but have the disadvantage that the estimator of $\Delta_{uv} = \tau_u - \tau_v$ depends on the observations from the other $k - 2$ treatments.

105. *Computational Difficulty*. Example 7.15 is a glaring illustration of the labor involved in computing W_{uv} when n is moderately large. It is necessary to obtain the median of $n(n + 1)/2$ Walsh averages, whereas the estimator Z_{uv} (7.37) is based on the median of only n differences. Thus, Doksum's contrast estimator is preferred to Lehmann's contrast estimator in terms of ease of computation. On the other hand, asymptotic efficiencies generally (but not always) favor Lehmann's contrast estimator. (See Section 7.16.)

Properties

1. *Standard Deviation of $\hat{\theta}$ (7.108).* For the asymptotic standard deviation of $\hat{\theta}$ (7.108), see Lehmann (1964).
2. *Asymptotic Normality*. See Lehmann (1964).
3. *Efficiency.* See Lehmann (1964) and Section 7.16.

Problems

127. Calculate the Lehmann estimator of the contrast $2\tau_N - \tau_A - \tau_R$ for the metronome data of Table 7.6. Compare with the Doksum estimator from Problem 46.

128. Give an example illustrating the incompatibility of the unadjusted estimators W_{uv} (7.105). (See Comment 104.)

129. Calculate the Lehmann estimators for the simple contrasts $\theta_1 = \tau_2 - \tau_1$, $\theta_2 = \tau_3 - \tau_1$, and $\theta_3 = \tau_3 - \tau_2$ for the CPK activity data in Table 7.3.

130. Estimate the contrast $3\tau_{\mathrm{ALC}} - \tau_{\mathrm{AL}} - \tau_{\mathrm{AC}} - \tau_{\mathrm{LC}}$ for the percentage consonants correctly identified data in Table 7.4.

131. Using the data of Table 7.4, obtain Lehmann's estimator of the simple contrast that represents the benefit from adding lip reading to audition in teaching severely hearing-impaired children. Compare with the Doksum estimator from Problem 50.

132. Compute Lehmann's estimator for all contrasts found to be of interest in Problem 45 for the maximum soil temperature data in Table 7.8.

133. Calculate Lehmann's estimator of the contrast $\tau_{\mathrm{rats}} - \tau_{\mathrm{cats}}$ for the Livesey EPT error score of Table 7.9.

7.16 EFFICIENCIES OF TWO-WAY LAYOUT PROCEDURES

We first consider the procedures of Sections 7.1–7.5, which are associated with the Friedman rank sums for the case of one observation per treatment–block combination (i.e., a randomized complete block design). The Pitman asymptotic relative efficiencies (for translation alternatives) of these procedures with respect to the corresponding normal theory counterparts are given by the expression

$$e_F = \left[\frac{k}{(k+1)}\right]\left[12\sigma_F^2\left\{\int_{-\infty}^{\infty} f^2(u)du\right\}^2\right], \tag{7.110}$$

where σ_F^2 is the variance of the common underlying (continuous) distribution F (7.1) and $f(.)$ is the probability density function corresponding to F. The parameter $\int_{-\infty}^{\infty} f^2(u)du$ is the area under the curve associated with $f^2(.)$, the square of the common probability density function. We note that e_F (7.110) is simply $k/(k+1)$ times the corresponding Pitman efficiencies in the one-sample, two-sample, and k-sample location settings (see Sections 3.11, 4.5, and 6.10).

In particular, the Pitman asymptotic relative efficiency of the Friedman test based on S (7.5) with respect to the normal theory two-way layout F test was found to be e_F (7.110) by van Elteren and Noether (1959). The asymptotic relative efficiency of the Page test for ordered alternatives, based on the statistic L (7.10), with respect to a suitable normal theory competitor was found by Hollander (1967a) to be e_F (7.110) as well. Furthermore, methods analogous to those of Sherman (1965) lead to expression (7.110) as the asymptotic relative efficiency of both the all-treatments two-sided and the treatments-versus-control one-sided multiple comparison procedures in Sections 7.3 and 7.4, respectively, with respect to the classical normal theory procedures based on sample means. Finally, Doksum (1967) obtained (7.110) as the asymptotic relative efficiency of

the estimator $\tilde{\theta}$ (7.40) with respect to the least-squares estimator $\bar{\theta} = \sum_{j=1}^{k} a_j X_{.j}$, where $X_{.j} = \sum_{i=1}^{n} X_{ij}/n$.

The efficiency e_F (7.110) is always greater than or equal to .576 and it can be infinite. Some values of e_F for various F and k combinations are given in Table 7.27.

We next turn to the procedures in Sections 7.6–7.8 that are designed for two-way layout data with zero or one observation per treatment–block combination. The Pitman asymptotic relative efficiency of the Durbin–Skillings–Mack test based on D (7.43) with respect to the standard normal theory procedure for a balanced incomplete block design was found to be e_F (7.110) by van Elteren and Noether (1959). Once again, methods analogous to those of Sherman (1965) lead to expression (7.110) as the asymptotic relative efficiency of the all-treatments two-sided multiple comparison procedures in Section 7.7. We do not know of any results for the asymptotic relative efficiencies of the general alternatives Skillings–Mack test in Section 7.8 for data from an arbitrary incomplete block design.

For the case of two-way layout data with at least one observation for every treatment–block combination, Mack and Skillings (1980) found that under certain conditions the asymptotic relative efficiency of their test for general alternatives based on the statistic MS (7.57) with respect to a suitable normal theory competitor is, once again, given by e_F (7.110). Combining their results with methods analogous to those of Sherman (1965) yields expression (7.110) as the asymptotic relative efficiency of the all-treatments two-sided multiple comparison procedures in Section 7.10, as well.

Finally, we turn to the procedures in Sections 7.11–7.15 which are associated with Wilcoxon signed ranks. The asymptotic relative efficiencies of Doksum's conservative test of Section 7.11, based on replacing λ_F by its upper bound $\frac{7}{24}$, are very close to those of a related test proposed by Doksum (1967) in which λ_F is estimated. The expression for the asymptotic relative efficiency e_F^* of the related test, relative to the normal theory $\mathcal{F}$-test, is given by the right-hand side of (2.12) in Doksum (1967). The parameter e_F^* is always greater than .864 and can be infinite. In Table 7.28, we provide values of e_F^* for normal, uniform and exponential distributions and various numbers (k) of treatments. Similarly, the efficiencies of Hollander's conservative test of Section 7.12, based on replacing λ_F by its upper bound $\frac{7}{24}$, are very close to those of a related test proposed by Hollander (1967a), in which λ_F is estimated. The expression for the asymptotic

Table 7.27 Values of e_F for Various Distributions and Numbers (k) of Treatments

k / Distribution	2	3	4	5	10	20	50	∞
				e_F				
Normal	0.637	0.716	0.764	0.796	0.868	0.909	0.936	0.955
Uniform	0.667	0.750	0.800	0.833	0.909	0.952	0.980	1.000
Double exponential	1.000	1.125	1.200	1.250	1.364	1.429	1.471	1.500

Table 7.28 Values of e_F^* for Various Distributions and Numbers (k) of Treatments

k / Distribution	2	3	4	5	10	20	50	∞
				e_F^*				
Normal	0.955	0.966	0.972	0.975	0.983	0.987	0.989	0.990
Uniform	0.889	0.894	0.897	0.899	0.902	0.904	0.905	0.906
Exponential	1.500	1.528	1.543	1.552	1.570	1.579	1.585	1.588

Table 7.29 Values of e_F^{**} for Various Distributions and Numbers (k) of Treatments

k	2	3	4	5	10	20	50	∞
Distribution				e_F^{**}				
Normal	0.955	0.963	0.969	0.972	0.980	0.985	0.988	0.990
Uniform	0.889	0.893	0.895	0.897	0.901	0.903	0.905	0.906
Exponential	1.500	1.521	1.534	1.543	1.563	1.575	1.583	1.588

relative efficiency e_F^{**} of this related test, with respect to a normal theory t-test for ordered alternatives, is given by the right-hand side of (4.6) of Hollander (1967a). The parameter e_F^{**} is always greater than .864 and can be infinite. In Table 7.29, we provide values of e_F^{**} for normal, uniform, and exponential distributions and various numbers (k) of treatments. The efficiencies in Table 7.29 are also close approximations to the efficiencies of the conservative multiple comparison procedures of Sections 7.13 and 7.14 with respect to normal theory competitors based on sample means. Lehmann (1964) obtained the asymptotic relative efficiency (for translation alternatives) of the contrast estimator (7.108) of Section 7.15 with respect to the least-squares estimator based on the sample means. The asymptotic relative efficiency is given by e_F^* (see Table 7.28).