*The Annals of Statistics*
1988, Vol. 16, No. 2, 751–758

# QUADRATIC LOSS OF ORDER RESTRICTED ESTIMATORS FOR TREATMENT MEANS WITH A CONTROL[1]

BY CHU-IN CHARLES LEE

*Memorial University of Newfoundland*

We consider an experiment which consists of $k$ treatment groups and a control group. Let the sample means $\overline{Y}_0, \overline{Y}_1, \dots, \overline{Y}_k$ be independent normal variates with expected values $\mu_0, \mu_1, \dots, \mu_k$ and with variances $\sigma^2/n_0, \sigma^2/n_1, \dots, \sigma^2/n_k$. Let $w_0, w_1, \dots, w_k$ be positive weights and let $\mu_0^*, \mu_1^*, \dots, \mu_k^*$ be the weighted least squares estimators subject to the constraints $\mu_0 \le \mu_i$, $i = 1, \dots, k$. We establish that for large $k$, $E(\mu_0^* - \mu_0)^2 > E(\overline{Y}_0 - \mu_0)^2$ when $w_i = n_i$, $i = 0, 1, \dots, k$. Under suitable conditions, we show that $E(\mu_i^* - \mu_i)^2 < E(\overline{Y}_i - \mu_i)^2$, $i = 0, 1, \dots, k$.

**1. Introduction.** We consider an experiment which consists of $k$ treatment groups and a control group. Let $Y_{ij}$, $j = 1, 2, \dots, n_i$, $i = 0, 1, \dots, k$, be independent normal variates with means $\mu_i$ and with a common variance $\sigma^2$, where $i = 0$ refers to the control group. If we are interested in determining which $\mu_i$ is significantly different from $\mu_0$, Dunnett's (1955) multiple range test may be applied. There are certain applications in the literature which exhibit the property that

$$\mu_0 \le \mu_i, \qquad i = 1, 2, \dots, k, \tag{1.1}$$

such as the blood cell counts in Dunnett (1955). The case in which all of the treatment means are no larger than the control mean can be treated by changing signs. The property (1.1) is known as the simple tree ordering [cf. Barlow, Bartholomew, Bremner and Brunk (1972)]. A likelihood ratio test for (1.1) against all alternatives can be found as a special case in Robertson and Wegman (1978). In this article, we shall assume that the simple tree ordering (1.1) is our prior knowledge. If we are interested in testing the hypothesis $\mu_0 = \mu_1 = \cdots = \mu_k$, it is to our advantage to restrict the parameter space accordingly [cf. Bartholomew (1961) and Robertson and Wright (1985)].

Let $\overline{Y}_0, \overline{Y}_1, \dots, \overline{Y}_k$ be the sample means for the $k+1$ groups. They are the unrestricted maximum likelihood estimators for $\mu_0, \mu_1, \dots, \mu_k$, they are unbiased but they may fail to satisfy (1.1). We are interested in utilizing the prior knowledge (1.1) to search for a better estimator with smaller mean square error pointwise than the usual estimator $(\overline{Y}_0, \overline{Y}_1, \dots, \overline{Y}_k)$. Let $w_0, w_1, \dots, w_k$ be positive weights and let $\mu^* = (\mu_0^*, \mu_1^*, \dots, \mu_k^*)$ be the weighted least squares estimator,

Received July 1986; revised September 1987.
[1]Research supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant A4541.
*AMS* 1980 *subject classifications*. Primary 62F10; secondary 62A10.
*Key words and phrases*. Isotonic regression, maximum likelihood estimator, order statistics, simple tree ordering.

i.e., $\mu^*$ minimizes

$$\sum_{i=0}^{k} (\overline{Y}_i - \mu_i)^2 w_i,$$

subject to the restriction (1.1). The weighted least squares estimator $\mu^*$ is known as the isotonic regression under the simple tree ordering [cf. Barlow, Bartholomew, Bremner and Brunk (1972)]. For the special case when the natural weights $w_i = n_i$, $i = 0, 1, \ldots, k$, are used, the isotonic regression $\mu^*$ is the maximum likelihood estimator subject to the restriction (1.1) and the notation $\hat{\mu}$ will be used.

The purposes of this article are three-fold. First, we establish that for large $k$,

$$E(\hat{\mu}_0 - \mu_0)^2 > E(\overline{Y}_0 - \mu_0)^2. \tag{1.2}$$

Second, we establish that

$$E(\mu_0^* - \mu_0)^2 < E(\overline{Y}_0 - \mu_0)^2 \tag{1.3}$$

for large values of $w_0$ (as compared with $w_1, \ldots, w_k$). Finally, we show that if $n_i \le n_0$, $i \ge 1$, then for any given positive weights $w_0, w_1, \ldots, w_k$,

$$E(\mu_i^* - \mu_i)^2 < E(\overline{Y}_i - \mu_i)^2, \tag{1.4}$$

where the condition $n_i \le n_0$ is necessary. It is quite common to have no less observations on the control than on the treatments. Ever since Lee (1981) showed that

$$E(\hat{\mu}_i - \mu_i)^2 < E(\overline{Y}_i - \mu_i)^2 \tag{1.5}$$

holds pointwise if (1.1) is replaced by $\mu_0 \le \mu_1 \le \cdots \le \mu_k$, there has been speculation that (1.5) might hold pointwise for all types of order restrictions. The reverse inequality (1.2) under the simple tree ordering (1.1) is the first counterexample in the literature. The inequality (1.3) indicates that the optimal weights are not necessarily the natural weights $n_i$.

The isotonic regression $\mu^*$ can be computed by the minimum lower set algorithm [cf. Brunk (1955)], the minimum violator algorithm [cf. Barlow, Bartholomew, Bremner and Brunk (1972)] and the min-max algorithm [cf. Lee (1983)]. The computation procedure is as follows. Let $\overline{Y}_{(1)} \le \overline{Y}_{(2)} \le \cdots \le \overline{Y}_{(k)}$ be the order statistics of $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_k$, let $w_{(i)}$ be the weights associated with $\overline{Y}_{(i)}$ and let

$$A_i = \sum_{j=0}^{i} w_{(j)} \overline{Y}_{(j)} \Big/ \sum_{j=0}^{i} w_{(j)},$$

where $w_{(0)} = w_0$ and $\overline{Y}_{(0)} = \overline{Y}_0$. We compute $A_0, A_1, \ldots, A_r$ successively until $A_r \le \overline{Y}_{(r+1)}$, $r < k$, or otherwise $A_{k-1} > \overline{Y}_{(k)}$, in which case we let $r = k$. By the max-min formula [cf. Barlow, Bartholomew, Bremner and Brunk (1972)], $\mu_0^* = A_r$

and it can be expressed by

$$\mu_0^* = \min_S \left( \sum_{i \in S} w_i \bar{Y}_i \Big/ \sum_{i \in S} w_i \right), \tag{1.6}$$

where $S$ is any subset of $\{0, 1, \ldots, k\}$ containing the element 0. The isotonic regression $\mu^*$ is obtained by setting $\mu_i^* = \max\{\mu_0^*, \bar{Y}_i\}$, $i = 1, \ldots, k$. It is clear that $\mu_0^* \le \bar{Y}_0$ and $\mu_i^* \ge \bar{Y}_i$, $i \ge 1$. Since these inequalities are strict with positive probabilities, the $\mu_i^*$ are biased. By Theorem 1.4 of Barlow, Bartholomew, Bremner and Brunk (1972), $\sum_{i=0}^k \mu_i^* w_i = \sum_{i=0}^k \bar{Y}_i w_i$, so the weighted sum is preserved. The isotonic regression $\mu^*$ is a continuous function of $\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_k$ and it is consistent.

**2. Counterexample.** Knowing that (1.1) holds doès not guarantee the mean square error reduction for the control group. Theorem 2.1 provides a counterexample by illustrating the unboundedness of the bias of $\hat{\mu}_0$.

**Theorem 2.1.** *If the means $\mu_0, \mu_1, \ldots, \mu_k$ and the sample sizes $n_0, n_1, \ldots, n_k$ are bounded, then for sufficiently large $k$,*

$$E(\hat{\mu}_0 - \mu_0)^2 > E(\bar{Y}_0 - \mu_0)^2.$$

**Proof.** Without loss of generality, we may assume $\mu_0 = 0$. It will be shown that

$$E\left(\hat{\mu}_0 I_{[\hat{\mu}_0 \le 0]}\right)^2 \to \infty \quad \text{as } k \to \infty.$$

By (1.6), $\hat{\mu}_0$ is decreasing in $k$ and symmetric in the $k$ treatment populations. Since a sample size must occur infinitely often in $n_1, n_2, \ldots,$ we may assume $n = n_1 = n_2 = \cdots$. Also $\hat{\mu}_0$ is increasing in $\mu_1, \mu_2, \ldots, \mu_k$, it suffices to assume $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ with $\mu$ the upper bound of the treatment means. Let $\bar{Y}_{(1)} = \min\{\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_k\}$ and $\tilde{\mu}_0 = (n_0\bar{Y}_0 + n\bar{Y}_{(1)})/(n_0 + n)$. By (1.6) again, $\hat{\mu}_0 \le \tilde{\mu}_0$. It suffices to show that $E(\tilde{\mu}_0 I_{[\tilde{\mu}_0 \le 0]})^2 \to \infty$ as $k \to \infty$. However, $\bar{Y}_{(1)} \to -\infty$ almost surely. Because $\tilde{\mu}_0$ is monotone in $k$, the monotone convergence theorem completes the proof. □

It is of interest to find the smallest integer $k$ satisfying (1.2). We consider the case $n_1 = n_2 = \cdots = n_k = n$ and $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$. Since $\hat{\mu}_0$ is increasing in $\mu$, under the extreme condition $\mu_0 = \mu$ we found by simulation that the $k$ can be as small as 5 provided $n_0 > 3.5n$. When $n_0 = n$, the smallest $k$ is 8 as shown in Table 1 where $\mu^* = \hat{\mu}$ when $w_i = 1$, $i = 0, 1, \ldots, k$.

**3. Mean square error at the control group.** By (1.6), $\mu_0^*$ is increasing in $w_0$. Therefore, the magnitude of the bias of $\mu_0^*$ is decreasing in $w_0$. Theorem 3.1 demonstrates that by increasing the weight $w_0$ the squared error reduction can be achieved. It does not require that $\mu$ satisfy (1.1).

TABLE 1

*Simulated mean square errors $E(\mu_i^* - \mu_i)^2$ under 100,000 iterations when $\overline{Y}_0, \overline{Y}_1, \ldots, \overline{Y}_k$ are independent normal variates with the same means and a common variance 1 and $w_1 = w_2 = \cdots = w_k = 1$*

| | $w_0 = 1$ | | $w_0 = k/3$ | |
|---|---|---|---|---|
| $k$ | $i = 0$ | $i \geq 1$ | $i = 0$ | $i \geq 1$ |
| 1 | 0.7500 | 0.7500 | 0.8139 | 0.8153 |
| 2 | 0.7308 | 0.7179 | 0.7805 | 0.7483 |
| 3 | 0.7662 | 0.7164 | 0.7662 | 0.7164 |
| 4 | 0.8191 | 0.7214 | 0.7540 | 0.6992 |
| 5 | 0.8817 | 0.7290 | 0.7520 | 0.6921 |
| 6 | 0.9387 | 0.7408 | 0.7483 | 0.6842 |
| 7 | 0.9982 | 0.7492 | 0.7458 | 0.6801 |
| 8 | 1.0553 | 0.7578 | 0.7445 | 0.6795 |
| 9 | 1.1086 | 0.7647 | 0.7434 | 0.6791 |
| 10 | 1.1695 | 0.7772 | 0.7424 | 0.6672 |
| 12 | 1.2632 | 0.7876 | 0.7306 | 0.6673 |
| 14 | 1.3610 | 0.7998 | 0.7289 | 0.6603 |
| 16 | 1.4509 | 0.8117 | 0.7340 | 0.6641 |
| 18 | 1.5300 | 0.8185 | 0.7315 | 0.6633 |
| 20 | 1.6072 | 0.8288 | 0.7352 | 0.6584 |
| 25 | 1.7769 | 0.8479 | 0.7324 | 0.6641 |

THEOREM 3.1. *Let the means $\mu_0, \mu_1, \ldots, \mu_k$, the sample sizes $n_0, n_1, \ldots, n_k$ and the positive weights $w_1, \ldots, w_k$ be fixed. There exists a positive real $W$ such that if $w_0 \geq W$, then*

$$E(\mu_0^* - \mu_0)^2 < E(\overline{Y}_0 - \mu_0)^2.$$

PROOF. Without loss of generality, we may assume $\mu_0 = 0$ and $\sigma^2/n_0 = 1$. Let $y_1 < y_2 < \cdots < y_k$ be observed values of $\overline{Y}_{(1)}, \ldots, \overline{Y}_{(k)}$, $w_{(i)}$ be the weight associated to $y_i$, $s_r = \sum_{i=1}^r w_{(i)}$, $a_r = \sum_{i=1}^r w_{(i)} y_i / s_r$ and $b_r = y_r + s_r(y_r - a_r)/w_0$ for $r = 1, \ldots, k$, and for convenience let $s_0 = 0$, $a_0 = a_1$, $b_0 = -\infty$ and $b_{k+1} = +\infty$. Then,

$$\mu_0^* = (w_0 \overline{Y}_0 + s_r a_r)/(w_0 + s_r) \quad \text{if } b_r < \overline{Y}_0 < b_{r+1}$$

for $r = 0, 1, \ldots, k$. Define a function $f(t)$ as $\mu_0^*$ as before with $\overline{Y}_0$ replaced by $t$ and $f(b_r) = y_r$, $r = 1, \ldots, k$. Then $f(t)$ is a strictly increasing continuous concave piecewise linear function with $f(t) = t$ if $t \leq y_1 = b_1$ and $(w_0 t + s_k y_1)/(w_0 + s_k) < f(t) \leq (w_0 t + s_1 y_1)/(w_0 + s_1)$ if $t > y_1$.

Suppose that $y_1 < 0$. There exists a positive real $\varepsilon$, $-s_1 y_1/w_0 \leq \varepsilon < -s_k y_1/w_0$, such that $f(\varepsilon) = 0$. Define a function $g(t)$ by

$$g(t) = t \quad \text{if } t \leq y_1$$

and

$$g(t) = -y_1(t - \varepsilon)/(\varepsilon - y_1) \quad \text{if } t > y_1.$$

Then $f(t)^2 \le g(t)^2$. Let $\Delta(t) = t^2 - f(t)^2$. Then

$$\int_{-\infty}^{\infty} \Delta(t)\phi(t)\,dt \ge \left\{-2\varepsilon y_1\overline{\Phi}(y_1) + \left(\tfrac{1}{2} - y_1^2\right)\varepsilon^2\right\}/(\varepsilon - y_1)^2$$

$$+\varepsilon^2\left\{y_1\phi(y_1) + \left(1 - y_1^2\right)\overline{\Phi}(y_1) - \tfrac{1}{2} + y_1^2\right\}/(\varepsilon - y_1)^2$$

$$> \tfrac{1}{2} - \tfrac{1}{2}y_1^2/(\varepsilon - y_1)^2 - y_1^2\varepsilon^2/(\varepsilon - y_1)^2$$

$$> \tfrac{1}{2} - \tfrac{1}{2}w_0^2/(w_0 + s_1)^2 - y_1^2 s_k^2/(w_0 + s_k)^2,$$

where $\phi(t)$ is the standard normal probability density function and $\overline{\Phi}(t)$ is the corresponding upper tail probability. The second term on the right side of the first inequality is positive. The last inequality is due to the observation that the functions $y_1^2/(t - y_1)^2$ and $t^2/(t - y_1)^2$ are decreasing and increasing, respectively, for $t > 0$ and the aforementioned two bounds for $\varepsilon$ are then evaluated respectively.

Suppose that $y_1 \ge 0$. It is trivial that the left-hand side of the preceding inequality is positive. Therefore,

$$E\left(\overline{Y}_0^2 - \mu_0^{*2}\right) = E\left[E\{\Delta(\overline{Y}_0)|\overline{Y}_1,\dots,\overline{Y}_k\}\right]$$

$$> \tfrac{1}{2}\left[1 - w_0^2/(w_0 + m)^2\right]P\left(\overline{Y}_{(1)} < 0\right) - s_k^2 E\left(\overline{Y}_{(1)}^2\right)/(w_0 + s_k)^2$$

$$> mP\left(\overline{Y}_{(1)} < 0\right)(w_0 - W)/(w_0 + s_k)^2,$$

with $W = s_k^2 E(\overline{Y}_{(1)}^2)/\{mP(\overline{Y}_{(1)} < 0)\}$ and $m = \min\{w_1,\dots,w_k\}$. The proof is complete. □

Theorem 3.1 indicates that the reduction of the mean square error $E(\mu_0^* - \mu_0)^2$ occurs for large $w_0$. In a simulation study (provided in Table 1) under the same means, the same sample sizes and $w_i = 1$, $i = 1,\dots,k$, we found that $w_0 = k/3$ is a suitable choice to significantly reduce the squared error for $\mu_0^*$ where $E(\overline{Y}_0 - \mu_0)^2$ is set to equal 1.

**4. Mean square errors at the treatment groups.** Brunk (1965) showed that for any given positive weights $w_0, w_1, \dots, w_k$,

$$\sum_{i=0}^{k} \left(\overline{Y}_i - \mu_i\right)^2 w_i \ge \sum_{i=0}^{k} \left(\overline{Y}_i - \mu_i^*\right)^2 w_i + \sum_{i=0}^{k} (\mu_i^* - \mu_i)^2 w_i,$$

provided that (1.1) holds. Thus the total weighted mean square error of the isotonic regression $\sum_{i=0}^k E(\mu_i^* - \mu_i)^2 w_i$ is strictly less than that of the usual estimator $\sum_{i=0}^k E(\overline{Y}_i - \mu_i)^2 w_i$. It follows that (1.4) holds for at least one $i$, $i = 0, 1, \dots, k$. Under suitable conditions, we shall show that (1.4) holds for all $i$. Theorems 4.1 and 4.2 do not require that $\mu$ satisfy (1.1).

THEOREM 4.1. *For a fixed index $i$, $1 \le i \le k$, if $\mu_i \ge \mu_0$ and $n_i \le n_0$, then*

$$E(\mu_i^* - \mu_i)^2 < E\left(\overline{Y}_i - \mu_i\right)^2$$

*for any given positive weights $w_0, w_1, \dots, w_k$.*

PROOF. By symmetry, it suffices to show the preceding inequality for $i = k$ when $\mu_k \geq \mu_0$ and $n_k \leq n_0$. Let $X = (\overline{Y}_{(1)}, \dots, \overline{Y}_{(k-1)})$, where $\overline{Y}_{(1)} \leq \overline{Y}_{(2)} \leq \cdots \leq \overline{Y}_{(k-1)}$ are the order statistics of $\overline{Y}_1, \overline{Y}_2, \dots, \overline{Y}_{k-1}$; let $Z = (\overline{Y}_0, \overline{Y}_k)$ and let

$$\Delta(Z) = (\overline{Y}_k - \mu_k)^2 - (\mu_k^* - \mu_k)^2.$$

By conditional expectation, it suffices to show that

$$E(\Delta(Z)|X = x) > 0 \tag{4.1}$$

for any given $x = (x_1, \dots, x_{k-1})$.

Without loss of generality, we may assume $\mu_0 = 0$. Consider a point $z = (\bar{y}_0, \bar{y}_k)$. If $\bar{y}_0 \leq \bar{y}_k$, then $\Delta(z) = 0$. We shall pair all other points $z_1 = (\bar{y}_{01}, \bar{y}_{k1})$ and $z_2 = (\bar{y}_{02}, \bar{y}_{k2})$ in the following manner:

$$\bar{y}_{01} - \bar{y}_{k1} = \bar{y}_{02} - \bar{y}_{k2} > 0 \tag{4.2}$$

and

$$(\bar{y}_{01} + \bar{y}_{k1}) - 2\mu_k = 2\mu_k - (\bar{y}_{02} + \bar{y}_{k2}) \geq 0. \tag{4.3}$$

Let $f(z)$ be the probability density function of the bivariate normal random vector $Z$. We shall establish the inequality

$$f(z_1)\Delta(z_1) + f(z_2)\Delta(z_2) \geq 0, \tag{4.4}$$

and that (4.4) holds strictly with positive probability. Consequently, (4.1) follows.

By (4.2) and (4.3), the conditions that $\mu_k \geq \mu_0$ and $n_k \leq n_0$ are sufficient for the inequality

$$f(z_1) \leq f(z_2). \tag{4.5}$$

Let $\mu_{01}^*$ and $\mu_{02}^*$ be the values of the isotonic regression $\mu^*$ at the index 0 evaluated at $z_1$ and $z_2$, respectively. If $\bar{y}_{k2} < \mu_{02}^*$, then $\mu_{k2}^* = \mu_{02}^* < \bar{y}_{02}$ and by (4.2) and (4.3),

$$\Delta(z_2) > 0. \tag{4.6}$$

Suppose that $\bar{y}_{k1} \geq \mu_{01}^*$. Then $\Delta(z_1) = 0$ and hence (4.4) holds. Suppose that $\bar{y}_{k1} < \mu_{01}^*$. For any nonnegative integer $t$, $t < k$, by (1.6) we have that

$$\begin{aligned}
&\left(w_0\bar{y}_{02} + \sum_{j=1}^{t} w_{(j)}x_j\right)\Bigg/\left(w_0 + \sum_{j=1}^{t} w_{(j)}\right) \\
&\quad = \left(w_0\bar{y}_{01} + \sum_{j=1}^{t} w_{(j)}x_j\right)\Bigg/\left(w_0 + \sum_{j=1}^{t} w_{(j)}\right) - w_0(\bar{y}_{01} - \bar{y}_{02})\Bigg/\left(w_0 + \sum_{j=1}^{t} w_{(j)}\right) \\
&\quad \geq \mu_{01}^* - (\bar{y}_{01} - \bar{y}_{02}) \\
&\quad > \bar{y}_{k1} - (\bar{y}_{01} - \bar{y}_{02}) \\
&\quad = \bar{y}_{k2},
\end{aligned}$$

where by convention $\sum_{j=1}^{t} w_{(j)} = \sum_{j=1}^{t} w_{(j)}x_j = 0$ if $t = 0$ and $w_{(j)}$ is the weight associated with $x_j$. Therefore, there is a nonnegative integer $r$, $r \le k - 1$, such that

$$\mu_{02}^* = \mu_{k2}^* = \left( w_0 \bar{y}_{02} + w_k \bar{y}_{k2} + \sum_{j=1}^{r} w_{(j)} x_j \right) \Big/ \left( w_0 + w_k + \sum_{j=1}^{r} w_{(j)} \right).$$

By (1.6) and (4.2), for the same integer $r$ we have that

$$\begin{aligned} \mu_{01}^* - \mu_{02}^* &\le \left( w_0 \bar{y}_{01} + w_k \bar{y}_{k1} + \sum_{j=1}^{r} w_{(j)} x_j \right) \Big/ \left( w_0 + w_k + \sum_{j=1}^{r} w_{(j)} \right) - \mu_{02}^* \\ &= (w_0 + w_k)(\bar{y}_{k1} - \bar{y}_{k2}) \Big/ \left( w_0 + w_k + \sum_{j=1}^{r} w_{(j)} \right) \\ &\le \bar{y}_{k1} - \bar{y}_{k2} \end{aligned}$$

and hence $\mu_{k1}^* - \bar{y}_{k1} \le \mu_{k2}^* - \bar{y}_{k2}$. It follows from (4.2) and (4.3) that

$$\Delta(z_1) + \Delta(z_2) > 0. \tag{4.7}$$

By (4.5)–(4.7), the inequality (4.4) holds strictly. This completes the proof. □

It is quite common to have no less observations on the control than on the treatments. Under that condition the inequality (1.4) holds for all $i$, $i = 1, \dots, k$, if (1.1) is satisfied. Furthermore, (1.4) holds for $i = 0$ as well if $w_0$ is sufficiently large. The condition $n_i \le n_0$ is necessary for Theorem 4.1 as illustrated in the example.

EXAMPLE. Let the sample means $\bar{Y}_0$ and $\bar{Y}_1$ be independent normal variates with expected values $\mu_0$ and $\mu_1$ and with variances $\sigma^2/n_0$ and $\sigma^2/n_1$. Let $\mu_0^*$ and $\mu_1^*$ be the weighted least squares estimators for $\mu_0$ and $\mu_1$ with positive weights $w_0$ and $w_1$ subject to the constraint $\mu_0 \le \mu_1$. Closed form expressions for the mean square errors $E(\mu_i^* - \mu_i)^2$, $i = 0, 1$, are available. When $\mu_0 = \mu_1$,

$$E(\mu_0^* - \mu_0)^2 = \sigma^2/n_0 + \sigma^2 w_1 [(n_0 - n_1) w_1 - 2 n_1 w_0] / [2 n_0 n_1 (w_0 + w_1)^2]$$

and

$$E(\mu_1^* - \mu_1)^2 = \sigma^2/n_1 + \sigma^2 w_0 [(n_1 - n_0) w_0 - 2 n_0 w_1] / [2 n_0 n_1 (w_0 + w_1)^2].$$

Therefore, if $n_0 > n_1$ and $w_0 < (n_0 - n) w_1 / 2 n_1$, it follows that we have $E(\mu_0^* - \mu_0)^2 > E(\bar{Y}_0 - \mu_0)^2$. Similarly, if $n_1 > n_0$ and $w_0 > 2 n_0 w_1 / (n_1 - n_0)$, then $E(\mu_1^* - \mu_1)^2 > E(\bar{Y}_1 - \mu_1)^2$. The total weighted mean square error reduction is $\sigma^2 2^{-1} (n_0^{-1} + n_1^{-1}) / (w_0^{-1} + w_1^{-1})$.

If one allows $w_0$ to vary, then the inequality (1.3) does not hold even when $k = 1$ as illustrated in the preceding example. The condition $n_i \le n_0$ may be relaxed according to Theorem 4.2.

THEOREM 4.2. *Let $\bar{\mu} = \sum_{i=0}^{k} n_i \mu_i / \sum_{i=0}^{k} n_i$. For a fixed index $i$, $1 \le i \le k$, if $\mu_i \ge \bar{\mu}$, then*

$$E(\hat{\mu}_i - \mu_i)^2 < E(\bar{Y}_i - \mu_i)^2.$$

PROOF. Let $\bar{\bar{Y}} = \sum_{i=0}^{k} n_i \bar{Y}_i / \sum_{i=0}^{k} n_i$. Recall that $\hat{\mu}_i = \bar{Y}_i$ if $\bar{Y}_i \geq \hat{\mu}_0$ and $\hat{\mu}_i = \hat{\mu}_0 \leq \bar{\bar{Y}}$ if $\bar{Y}_i < \hat{\mu}_0$. Therefore,

$$\begin{aligned} E(\bar{Y}_i - \mu_i)^2 - E(\hat{\mu}_i - \mu_i)^2 &= E(\bar{Y}_i - \hat{\mu}_i)(\bar{Y}_i + \hat{\mu}_i - 2\mu_i) I_{[\bar{Y}_i < \hat{\mu}_0]} \\ &= E(\bar{Y}_i - \hat{\mu}_0)(\bar{Y}_i + \hat{\mu}_0 - 2\bar{\bar{Y}}) I_{[\bar{Y}_i < \hat{\mu}_0]} \\ &\quad + 2E(\bar{Y}_i - \hat{\mu}_0)(\bar{\bar{Y}} - \mu_i) I_{[\bar{Y}_i < \hat{\mu}_0]} \\ &> 2E(\bar{Y}_i - \hat{\mu}_0)(\bar{\bar{Y}} - \mu_i) I_{[\bar{Y}_i < \hat{\mu}_0]} \\ &= 2E(\bar{\bar{Y}} - \mu_i) E(\bar{Y}_i - \hat{\mu}_0) I_{[\bar{Y}_i < \hat{\mu}_0]} \geq 0, \end{aligned}$$

where the last identity is due to the fact that $\bar{\bar{Y}}$ is independent of $\bar{Y}_i - \hat{\mu}_0$. This completes the proof. □

**Acknowledgments.** The author is very grateful to Professor Tim Robertson for suggesting this investigation. The author thanks the two referees and Professor T. K. Mak for valuable comments. The revised proof for Theorem 2.1 is due to one of the referees.

## REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, New York.

BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives (with discussion). *J. Roy. Statist. Soc. Ser. B* **23** 239–281.

BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

BRUNK, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

LEE, C. C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.* **9** 686–688.

LEE, C. C. (1983). The min-max algorithm and isotonic regression. *Ann. Statist.* **11** 467–477.

ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for ordered restrictions in exponential families. *Ann. Statist.* **6** 485–505.

ROBERTSON, T. and WRIGHT, F. T. (1985). One-sided comparisons for treatments with a control. *Canad. J. Statist.* **13** 109–122.

DEPARTMENT OF MATHEMATICS AND STATISTICS
MEMORIAL UNIVERSITY OF NEWFOUNDLAND
ST. JOHN'S, NEWFOUNDLAND A1C 5S7
CANADA