한국통계학회 논문집 제1권 1호
1994년 12월 pp. 57-74

A Common Mean Estimation Problem of P-Normal Populations [1]

Seung-soo Lee, Kwan-young Kim[2]

Abstract

This paper deals with the estimation problems of a common mean of several independent normal populations with unknown variances, based on random samples of equal size. The authors suggest a promising approach and a new estimator to improve Graybill-Deal estimator further. By Monte Carlo simulation study, the efficiency of new estimator is compared with that of Graybill-Deal estimator.

1. Introduction

Let $X_{i1}, \cdots, X_{iN}$ be independent random samples from a normal distribution with a common mean μ and unknown variance σ_i^2, $i=1,\cdots,P$, respectively. If each σ_i^2 is known, the best (minimum variance unbiased) estimator of common mean μ is

$$\hat{\mu}^* = (\sum_{i=1}^{P} \overline{X_i}/\sigma_i^2)/(\sum_{i=1}^{P} 1/\sigma_i^2),$$

where $\overline{X_i} = (\sum_{j=1}^{N} X_{ij})/N$. In the case of unknown variances, Graybill and Deal(1959) suggested

$$\hat{\mu}_G = (\sum_{i=1}^{P} \overline{X_i}/S_i^2)/(\sum_{i=1}^{P} 1/S_i^2),$$

as an estimator of μ, where $S_i^2 = \sum_{j=1}^{N}(X_{ij} - \overline{X_i})^2$. They proved that $\hat{\mu}_G$ dominate $\overline{X_i}$ in variance for P=2, if N>10.

For the further references, see Zacks(1966), Mehta and Gurland(1969), Cohen and Sackrowitz(1974), Bhattacharya(1978), Sinha and Mouqadem(1982), and etc.

For $P \geq 2$, Norwood and Hinkelmann(1977) gave a necessary and sufficient condition that $\hat{\mu}_G$ dominate $\overline{X_i}$, $i=1,\cdots,P$. There are only a few papers for $P \geq 2$. Confer Brown and

---

1) This paper was supported by Institute of Basic Science at Kangwon National University, 1993.
2) Department of Statistics, Kangwon National University, 192-1 hyoja 2-dong, chuncheon, KOREA 200-701

Cohen(1974), Shinozaki(1978) and Rao(1980).

In this paper, we suggest a class of estimators of common mean which can dominate Graybill-Deal's estimator in efficiency. We present a class of estimators of μ which have the form

$$\hat{\mu}_{NEW} = (1-\lambda)\ \hat{\mu}_G + \lambda\bar{\mu}, \tag{1.1}$$

where grand mean $\bar{\mu} = \sum_{i=1}^{P} \overline{X_i}/P$ and $0<\lambda<1$. Note that if $\lambda=0$, then $\hat{\mu}_{NEW} = \hat{\mu}_G$ and if $\lambda=1$, then $\hat{\mu}_{NEW} = \bar{\mu}$. λ plays a role of shrinkage variable between $\hat{\mu}_G$ and $\bar{\mu}$ by a degree of equality of variances.

The motivation for deriving the above estimator is closely related to the problem of Stein in estimation of multidimensional parameters.

$\underline{1}$ denotes the (P $\times$ 1) column vector with each components 1 and $\underline{V}$ denotes the (P $\times$ P) nonsingular positive definite matrix with the elements $\sigma_1^2, \cdot\cdot\cdot, \sigma_P^2$ in its main diagonal positions and zeros in all other locations. Transpose of a vector or matrix $\underline{A}$ is written by $\underline{A}^T$. The inverse matrix of a nonsingular positive definite matrix B is expressed $\underline{B}^{-1}$. (a_{ij}) will denote the matrix with elements a_{ij}.

Let $\underline{X}_j, j=1, \cdot\cdot\cdot, N$ be iid P- dimensional randon vector having multivariate normal distribution with mean vector $\mu\underline{1}$(where μ is a scalar) and variance covariance matrix $\underline{V}$.

Our problem is the case when $\underline{V}$ is diagonal , but now we consider a general $\underline{V}$.

When $\underline{V}$ is known the best estimator of μ is obtained by

$$\hat{\mu}^* = (\underline{1}^T\underline{V}^{-1}\overline{\underline{X}})/(\underline{1}^T\underline{V}^{-1}\underline{1})$$

where $\overline{\underline{X}} = (\sum_{j=1}^{N}\underline{X}_j)/N$, and when $\underline{V}$ is unknown, we may have

$$\hat{\mu}_M = (\underline{1}^T\underline{S}^{-1}\overline{\underline{X}})/(\underline{1}^T\underline{S}^{-1}\underline{1})$$

where S denotes the matrix of sum of squares and cross products of $\underline{X}_j$s. We can rewrite $\hat{\mu}_G$ as

$$\hat{\mu}_G = (\underline{1}^T\underline{D}^{-1}\overline{\underline{X}})/(\underline{1}^T\underline{D}^{-1}\underline{1})$$

where $\underline{D}$ denotes the (P $\times$ P) diagonal matrix with elements $S_1^2, \cdot\cdot\cdot, S_P^2$ in its main diagonal positions and zeros in all other locations. In Section 2, we will show that $\hat{\mu}^*$can be expressed in the alternative form as

$$\hat{\mu}^* = \bar{\mu} + \underline{B}^T\overline{\underline{Y_2}}$$

where $\overline{\mu}=\underline{1}^T\underline{\overline{X}}/P$ and $\underline{\overline{Y_2}}$ is P-1 dimensional random vector and $\underline{B}$ is the regression coefficient on $\underline{\overline{Y_2}}$. $\hat{\mu}_M$ is also expressed as

$$\hat{\mu}_M=\overline{\mu}+\underline{\widehat{B^T}\overline{Y_2}}$$

when $\hat{B}$ is estimated by the least squares method. When P-1 ≥ 3 (i.e. P ≥ 4), it was shown by Baranchik(1973) that $\underline{\hat{B}}$ can be improved quite analogously as the case of the mean vector of the multivariate normal distribution by shrinking. That is, $\hat{\mu}_M$ is dominated by an estimator of the form

$$\hat{\mu}_B=\overline{\mu}+\underline{\widehat{B^{*T}}}\ \underline{\overline{Y_2}}.$$

In our case it is expected that Baranchik's estimator can still be improved by making use of the condition that $\underline{V}$ is diagonal, and again we get the estimation of the type (1.1).

More detailed discussion of the derivation is given in the next section.

In Section 3, we evaluate the variance of $\hat{\mu}_{NEW}$, $Var(\hat{\mu}_{NEW})$. For P ≥ 2, in case of equal variances, we prove that the efficiency of $\hat{\mu}_{NEW}$ is larger than that of $\hat{\mu}_G$.

For P ≥ 3, it is very difficulty to evaluate $Var(\hat{\mu}_{NEW})$ analytically (e.g. $Var(\hat{\mu}_G)$ has not been evaluated exactly in literature.). Therefore in order to examine $Var(\hat{\mu}_{NEW})$, we carry out Monte Carlo simulation and compare efficiencies of our estimator with that of Graybill-Deal estimator.

Conclusions are summarized in Section 5.

2. Rationale for a new unbiased estimator

In this Section, we give a rationale for choice of new estimator $\hat{\mu}_{NEW}$. To begin with, we consider the following transfomation. That is, we define

$$Y_{1j}\equiv \sum_{i=1}^{P} X_{ij}/P \tag{2.1}$$

and

$$Y_{ij}\equiv \sum_{k=1}^{P} a_{ik}X_{kj},\ i=2,\cdots,P,\ j=1,\cdots,N \tag{2.2}$$

where

$$\sum_{k=1}^{P} a_{ik} = 0$$

$$\sum_{k=1}^{P} a_{ik}a_{jk} = 1 \ , \ i=j$$
$$\qquad\qquad 0 \ , \ i \neq j.$$

From (2.1) and (2.2) we also define

$$\overline{Y_1} = \sum_{j=1}^{N} Y_{1j}/N = (\underline{1}^T \underline{\overline{X}})/P$$

and

$$\underline{\overline{Y_2}} = \frac{1}{N}\begin{pmatrix} \sum_{j=1}^{N} Y_{2j} \\ \sum_{j=1}^{N} Y_{3j} \\ \vdots \\ \sum_{j=1}^{N} Y_{Pj} \end{pmatrix} = \begin{pmatrix} \overline{Y_2} \\ \overline{Y_3} \\ \vdots \\ \overline{Y_P} \end{pmatrix} = \underline{A}^T \underline{\overline{X}}$$

where $\underline{A}^T = (a_{ij})$ $i=2, \ldots, P$ and $j=1,\ldots, P$. Then we define as follows

$$\underline{\overline{Y}} \equiv \begin{pmatrix} \overline{Y_1} \\ \underline{\overline{Y}}_2 \end{pmatrix} = \begin{pmatrix} \underline{1}^T \underline{\overline{X}}/P \\ \underline{A}^T \underline{\overline{X}} \end{pmatrix} = \begin{pmatrix} \underline{1}^T/P \\ \underline{A}^T \end{pmatrix} \underline{\overline{X}}.$$

Since

$$\begin{pmatrix} \underline{1}^T/P \\ \underline{A}^T \end{pmatrix} (\underline{1} \ \ \underline{A}) = \begin{pmatrix} 1 & 0 \\ 0 & \underline{I} \end{pmatrix} = \underline{I}_{(P \times P)}, \tag{2.3}$$

we get

$$\underline{\overline{X}} = (\underline{1} \ \ \underline{A})\underline{\overline{Y}} = \underline{1}\overline{Y}_1 + \underline{A}\,\underline{\overline{Y}}_2.$$

From the above we can rewrite $\hat{\mu}^*$ as follows

$$\begin{aligned} \hat{\mu}^* &= [(\underline{1}^T \underline{V}^{-1} \underline{1})\overline{Y_1} + (\underline{1}^T \underline{V}^{-1} \underline{A})\underline{\overline{Y_2}}]/(\underline{1}^T \underline{V}^{-1} \underline{1}) \\ &= \overline{Y_1} + (\underline{1}^T \underline{V}^{-1} \underline{1})^{-1} (\underline{1}^T \underline{V}^{-1} \underline{A}) \underline{\overline{Y}}_2 \ . \end{aligned}$$

Defining $(1 \times (P-1))$ vector

$$\underline{B}^T = (\underline{1}^T \underline{V}^{-1} \underline{1})^{-1} (\underline{1}^T \underline{V}^{-1} \underline{A})$$

we can write $\hat{\mu}^* = \overline{Y_1} + \sum_{i=2}^{P} \beta_i \overline{Y_i}$, where $\underline{B}^T = (\beta_2, \cdots, \beta_P)$. Note that with unknown $\underline{V}$, the problem of estimating μ is equivalent to that of $\underline{B}$.

From now on, in order to estimate $\underline{B}$, we'll consider the conditional distribution of Y_1, given $\underline{Y}_2 = (Y_2, \cdots Y_P)^T$ where $Y_1 \equiv \sum_{i=1}^{P} X_i / P$ and $Y_i \equiv \sum_{k=1}^{P} a_{ik} X_k$, $i = 2, \cdots, P$. (a_{ik}) equals in (2.2). Then the conditional distribution has mean

$$\mu - \underline{B}^{*T} \underline{Y}_2$$

and variance

$$\sigma^2_{Y_1|\underline{Y}_2} = V_{Y_1Y_1} - \underline{V}_{Y_1\underline{Y}_2} \underline{V}^{-1}_{\underline{Y}_2\underline{Y}_2} \underline{V}_{\underline{Y}_2Y_1}, \tag{2.4}$$

where $\underline{B}^* = -\underline{V}^{-1}_{\underline{Y}_2\underline{Y}_2} \underline{V}_{\underline{Y}_2Y_1}$, $V_{Y_1Y_1} = (\underline{1}^T \underline{V} \underline{1})/P^2$, $\underline{V}_{\underline{Y}_2\underline{Y}_2} = \underline{A}^T \underline{V} \underline{A}$, $\underline{V}_{\underline{Y}_2Y_1} = (\underline{A}^T \underline{V} \underline{1})/P$, $\underline{V}_{Y_1\underline{Y}_2} = (\underline{1}^T \underline{V} \underline{A})/P$.

<u>Lemma 2.1</u>. $\underline{B} = \underline{B}^*$.

<u>proof</u> . We note that

$$\begin{pmatrix} \underline{1}^T \\ \underline{A}^T \end{pmatrix} \underline{V}^{-1} (\underline{1}/P \;\; \underline{A}) = \begin{pmatrix} (\underline{1}^T \underline{V}^{-1} \underline{1})/P & \underline{1}^T \underline{V}^{-1} \underline{A} \\ (\underline{A}^T \underline{V}^{-1} \underline{1})/P & \underline{A}^T \underline{V}^{-1} \underline{A} \end{pmatrix}. \tag{2.5}$$

From (2.3), the inverse matrix of (2.5) becomes

$$\left(\begin{pmatrix} \underline{1}^T \\ \underline{A}^T \end{pmatrix} \underline{V} (\underline{1}/P \;\; \underline{A}) \right) = \left(\begin{pmatrix} (\underline{1}^T \underline{V} \underline{1})/P & \underline{1}^T \underline{V} \underline{A} \\ (\underline{A}^T \underline{V} \underline{1})/P & \underline{A}^T \underline{V} \underline{A} \end{pmatrix} \right). \tag{2.6}$$

By simple calculations of matrix, we get

$$(\underline{A}^T \underline{V}^{-1} \underline{1})/P = -(\underline{A}^T \underline{V} \underline{A})^{-1} (\underline{A}^T \underline{V} \underline{1}) (\underline{1}^T \underline{V}^{-1} \underline{1})/P^2.$$

Accordingly, since

$$(\underline{A}^T \underline{V}^{-1} \underline{1})(\underline{1}^T \underline{V}^{-1} \underline{1})^{-1} = -(\underline{A}^T \underline{V} \underline{A})^{-1} (\underline{A}^T \underline{V} \underline{1})/P,$$

we get

$$\underline{B} = \underline{B}^* . \qquad ///$$

From the conditional distribution of Y_1 given $\underline{Y_2}$, it is easy to show that the maximum liklihood estimators of μ and $\underline{B}$ are, respectively,

$$\hat{\mu}_M = \overline{Y_1} + \hat{\underline{B}}^T \overline{\underline{Y_2}}$$

and

$$\underline{\hat{B}} = -\underline{S}^{-1}_{\underline{Y_2}\underline{Y_2}}\underline{S}_{\underline{Y_2}Y_1},$$

where $\underline{S}^{-1}_{Y_2\underline{Y_2}} = \underline{A}^T\underline{S}\underline{A}$ and $\underline{S}_{\underline{Y_2}Y_1} = (\underline{A}^T\underline{S}\underline{1})/P$. S denotes the matrix of sums of squares and cross products

$$\underline{S} = \sum_{j=1}^{N}(\underline{X}_j - \overline{\underline{X}})(\underline{X}_j - \overline{\underline{X}})^T.$$

From (2.4), (2.5) and (2.6), by similar procedures, we get

$$\sigma^2_{Y_1|\underline{Y_2}} = (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}. \tag{2.7}$$

By the well-known property in regression analysis, given $\underline{Y_2}$, $\underline{\hat{B}}$ has (P-1) dimensional normal distribution with mean $\underline{B}$ and variance $\underline{S}^{-1}_{\underline{Y_2}\underline{Y_2}}\sigma^2_{Y_1|\underline{Y_2}}$. Hence we obtain the following lemma.

Lemma 2.2 . The quantity

$$\hat{\mu}_M = \overline{Y_1} + \underline{\hat{B}}^T\overline{\underline{Y_2}}$$

is an unbiased estimator of μ and has the variance

$$Var(\hat{\mu}_M) = (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}[1+(P-1)/(N-P-1)]/N.$$

Proof. Unbiasedness is directly obtained, since

$$\begin{aligned} E(\hat{\mu}_M) &= E_{\underline{Y_2}}E_{Y_1|\underline{Y_2}}(\hat{\mu}_M|\underline{Y_2}) \\ &= E_{\underline{Y_2}}(\mu + \underline{B}^T\overline{\underline{Y_2}}) = \mu . \end{aligned}$$

For the variance, first we note that

$$\begin{aligned} Var(\hat{\mu}_M) &= Var(\hat{\mu}^* - \hat{\mu}^* + \hat{\mu}_M) \\ &= Var(\hat{\mu}^*) + Var(\hat{\mu}^* - \hat{\mu}_M), \end{aligned}$$

because $\hat{\mu}^*$ is the best. Accordingly

$$\begin{aligned} Var(\hat{\mu}_M) &= (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}/N + Var[(\underline{\hat{B}} - \underline{B})^T\overline{\underline{Y_2}}] \\ &= (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}/N + E[(\underline{\hat{B}} - \underline{B})^T\underline{V}_{\underline{Y_2}\underline{Y_2}}(\underline{\hat{B}} - \underline{B})]/N \\ &= (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}/N + E_{\underline{Y_2}}[tr(\underline{V}_{\underline{Y_2}\underline{Y_2}}\underline{S}^{-1}_{\underline{Y_2}\underline{Y_2}}\sigma^2_{Y_1|\underline{Y_2}})]/N , \end{aligned} \tag{2.8}$$

where "tr" denotes the trace of matrix. Then using the property of Whishart distribution and (2.7), we obtain that

$$Var(\hat{\mu}_M) = (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}[1+(P-1)/(N-P-1)]/N . \quad ///$$

Seeing the proof of lemma 2.2, we know that goodness of an estimator, $\hat{\mu}_o$, depends on

(2.8), that is,

$$E[(\hat{\underline{B}}_o-\underline{B})^T \underline{V_{Y_2Y_2}}(\hat{\underline{B}}_o-\underline{B})] \tag{2.9}$$

Therefore, we can use (2.9) as the risk function in estimating $\underline{B}$. In other words, the best estimator of μ is the one minimizing (2.9). Let's consider the following estimator of $\underline{B}$,

$$\hat{\underline{B}}^* = [1-c\,\hat{\sigma}^2/(\hat{\underline{B}}^T \underline{S_{Y_2Y_2}}\hat{\underline{B}})]\hat{\underline{B}}$$

where

$$\hat{\sigma}^2 = S^2_{Y_1} - \underline{S}_{Y_1Y_2}\underline{S^{-1}_{Y_2Y_2}}\underline{S}_{Y_2Y_1}$$

$$= (\underline{1}^T \underline{S}^{-1}\underline{1})^{-1}$$

and

$$c \in \{ 0 \ , \ 2(P-3)/(N-P+3) \ \}.$$

Baranchik(1973) has proved that

$$E[(\hat{\underline{B}}^*-\underline{B})^T \underline{V_{Y_2Y_2}}(\hat{\underline{B}}^*-\underline{B})] < E[(\hat{\underline{B}}-\underline{B})^T \underline{V_{Y_2Y_2}}(\hat{\underline{B}}-\underline{B})],$$

if $P \geq 4$ and $N \geq P+1$. Namely we can improve $\hat{\mu}_M$ by adopting

$$\hat{\mu}_B = \overline{Y_1} + \hat{B}^{*T}\underline{\overline{Y}}_2.$$

Considering the circumstances mentioned above, we have known that $\hat{\mu}_M$ is dominated by $\hat{\mu}_B$ in variance. Here, if we employ independence property among populations it is natural to expect improvements on $\hat{\mu}_B$. For that purposes we'll take up "Rao-Blackwellizing" on $\hat{\underline{B}}^*$.[Bickel and Doksum(1977, 121p)]

By such a procedure in the proof of lemma 2.1, since we can rewrite $\hat{\underline{B}}$ as follows

$$\hat{\underline{B}} = (\underline{1}^T\underline{S}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{S}^{-1}\underline{1}),$$

we get

$$\widehat{\underline{B}^*} = \{1-cP^2/\{(\underline{1}^T\underline{S}\,\underline{1})(\underline{1}^T\underline{S}^{-1}\underline{1})-P^2\}\}\hat{\underline{B}}. \tag{2.10}$$

As the variance covariance matrix $\underline{V}$ is the diagonal matrix, its sufficient statistic is $\underline{D}$. Hence, we'll define $\hat{\underline{B}}^{**}$ as follows,

$$\hat{\underline{B}}^{**} = E[\hat{\underline{B}}^*|\underline{D}]. \tag{2.11}$$

In order to find $\hat{\underline{B}}^{**}$, we rewrite the relevant statistics as follows,

$$\underline{1}^T\underline{S}\,\underline{1} = \underline{1}^T\underline{D}^{1/2}\underline{R}\,\underline{D}^{1/2}\underline{1} \equiv \underline{a}^T\underline{R}\,\underline{a}$$

and

$$\underline{1}^T\underline{S}^{-1}\underline{1} = \underline{1}^T\underline{D}^{-1/2}\underline{R}^{-1}\underline{D}^{-1/2}\underline{1} \equiv \underline{b}^T\underline{R}^{-1}\underline{b},$$

where $\underline{R}$ is a sample correlation matrix of $\underline{X}$ and $\underline{D} = \underline{D}^{1/2}\underline{D}^{1/2}$ and $\underline{D}^{-1} = \underline{D}^{-1/2}\underline{D}^{-1/2}$. Then substituting them in (2.10), we get

$$\hat{\underline{B}}^* = \{1 - cP^2/\{(\underline{a}^T\underline{Ra})(\underline{b}^T\underline{R}^{-1}\underline{b}) - P^2\}\}(\underline{G}^T\underline{R}^{-1}\underline{b})(\underline{b}^T\underline{R}^{-1}\underline{b})^{-1},$$

where $\underline{G}^T = \underline{A}^T\underline{D}^{-1/2}$.

But since we can not calculate (2.11), exactly, we'll utilize an apprximation of $\hat{\underline{B}}^*$ by Taylor's expansions. First we divide $\underline{R}$ as follows

$$\underline{R} = \underline{I} + \underline{\Delta R}$$

where $\underline{I}$ is the (P × P) identity matrix and ΔR is $\underline{R} - \underline{I}$. Using the Taylor's expansions, we acquire

$$\underline{a}^T\underline{Ra} = \underline{a}^T(\underline{I} + \underline{\Delta R})\underline{a} = \underline{a}^T\underline{a} + \underline{a}^T\underline{\Delta R}\,\underline{a},$$

$$\underline{b}^T\underline{R}^{-1}\underline{b} = \underline{b}^T(\underline{I} + \underline{\Delta R})^{-1}\underline{b} \approx \underline{b}^T\underline{b} - \underline{b}^T\underline{\Delta R}\,\underline{b} + \underline{b}^T\underline{\Delta^2 R}\,\underline{b},$$

$$\underline{G}^T\underline{R}^{-1}\underline{b} = \underline{G}^T(\underline{I} + \underline{\Delta R})^{-1}\underline{b} \approx \underline{G}^T\underline{b} - \underline{G}^T\underline{\Delta Rb} + \underline{G}^T\underline{\Delta^2 R\,b},$$

$$(\underline{a}^T\underline{Ra})(\underline{b}^T\underline{R}^{-1}\underline{b}) \approx \underline{a}^T\underline{a}\,\underline{b}^T\underline{b} + (\underline{b}^T\underline{b})(\underline{a}^T\underline{\Delta R}\,\underline{a}) - (\underline{a}^T\underline{a})(\underline{b}^T\underline{R}^{-1}\underline{b})$$
$$+ (\underline{a}^T\underline{a})(\underline{b}^T\underline{\Delta^2 Rb}) - (\underline{a}^T\underline{\Delta R}\,\underline{a})(\underline{b}^T\underline{\Delta R}\,\underline{b}),$$

$$cP^2/\{(\underline{a}^T\underline{Ra})(\underline{b}^T\underline{R}^{-1}\underline{b}) - P^2\} \approx \lambda_1\{1 - (\underline{b}^T\underline{b})(\underline{a}^T\underline{\Delta Ra})/K$$
$$+ (\underline{a}^T\underline{a})(\underline{b}^T\underline{\Delta R}\,\underline{b})/K - (\underline{a}^T\underline{a})(\underline{b}^T\underline{\Delta^2 R}\,\underline{b})/K + (\underline{a}^T\underline{\Delta Ra})(\underline{b}^T\underline{\Delta Rb})$$
$$+ (\underline{b}^T\underline{b})^2(\underline{a}^T\underline{\Delta Ra})^2/K^2 + (\underline{a}^T\underline{a})^2(\underline{b}^T\underline{\Delta Rb})^2/K^2$$
$$- 2(\underline{a}^T\underline{a})(\underline{b}^T\underline{b})(\underline{a}^T\underline{\Delta Ra})(\underline{b}^T\underline{\Delta R b})/K^2,$$

where $\lambda_1 = cP^2/\{(\underline{a}^T\underline{a})(\underline{b}^T\underline{b}) - P^2\}$, $K = (\underline{a}^T\underline{a})(\underline{b}^T\underline{b}) - P^2$ and

$$(\underline{b}^T\underline{R}^{-1}\underline{b})^{-1}(\underline{G}^T\underline{R}^{-1}\underline{b}) \approx (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{b} - \underline{G}^T\underline{\Delta Rb} + \underline{G}^T\underline{\Delta^2 R\,b})$$
$$\{1 + (\underline{b}^T\underline{b})^{-1}(\underline{b}^T\underline{\Delta Rb}) - (\underline{b}^T\underline{b})^{-1}(\underline{b}^T\underline{\Delta^2 R\,b}) + (\underline{b}^T\underline{b})^{-2}(\underline{b}^T\underline{\Delta Rb})^2\},$$

where $\approx$ denotes the near equal. In these approximate expressions we did not write the error terms, because it is very tedious and complicated and we don't need them. Using those expressions, we can approximate $\hat{\underline{B}}^*$ as follows

$$\hat{\underline{B}}^* \approx (1 - \lambda_1)(\underline{b}^T\underline{b})^{-1}\{\underline{G}^T\underline{b} - \underline{G}^T\underline{\Delta Rb} + \underline{G}^T\underline{\Delta^2 R\,b}$$
$$+ (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{b})(\underline{b}^T\underline{\Delta Rb}) + (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{\Delta R b})(\underline{b}^T\underline{\Delta R b})$$
$$- (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{b})(\underline{b}^T\underline{\Delta^2 R\,b}) + (\underline{b}^T\underline{b})^{-2}(\underline{G}^T\underline{b})(\underline{b}^T\underline{\Delta R b})^2\}$$
$$+ (\underline{b}^T\underline{b})^{-1}\{\lambda_1(\underline{b}^T\underline{b})(\underline{a}^T\underline{\Delta Ra})/K - \lambda_1(\underline{a}^T\underline{a})(\underline{b}^T\underline{\Delta Rb})/K\}$$
$$\{\underline{G}^T\underline{b} - \underline{G}^T\underline{\Delta Rb} + (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{b})(\underline{b}^T\underline{\Delta R b})\}$$
$$+ (\underline{b}^T\underline{b})^{-1}(\underline{G}^T\underline{b})\{\lambda_1(\underline{a}^T\underline{a})(\underline{b}^T\underline{\Delta^2 R\,b})/K - \lambda_1(\underline{a}^T\underline{\Delta Ra})(\underline{b}^T\underline{\Delta R\,b})/K$$
$$- \lambda_1(\underline{b}^T\underline{b})^2(\underline{a}^T\underline{\Delta Ra})^2/K^2 - \lambda_1(\underline{a}^T\underline{a})^2(\underline{b}^T\underline{\Delta Rb})^2/K^2$$
$$+ 2\lambda_1(\underline{a}^T\underline{a})^2(\underline{b}^T\underline{b})^2(\underline{a}^T\underline{\Delta Ra})(\underline{b}^T\underline{\Delta Rb})/K^2\}.$$

By taking the conditional expectation on the above equation, we'll obtain the approximate variates of $\underline{\hat{B}}^{**}$ as follows

$$\underline{\hat{B}}^{**} \approx (1-\lambda_1)\{(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) - 2(\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}\{(\underline{1}^T\underline{D}^{-1}\underline{1})(\underline{A}^T\underline{D}^{-1}\underline{1}) - \underline{A}^T\underline{D}^{-2}\underline{1}\}/(N-1)$$

$$+ 2(\underline{1}^T\underline{D}^{-1}\underline{1})^{-3}(\underline{A}^T\underline{D}^{-1}\underline{1})\{(\underline{1}^T\underline{D}^{-1}\underline{1})^2 - \underline{1}^T\underline{D}^{-2}\underline{1}\}/(N-1)\}$$

$$-2\lambda_1\{(P-1)\underline{A}^T\underline{1} - (\underline{1}^T\underline{D}\underline{1})(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}\{(\underline{1}^T\underline{D}^{-1}\underline{1})(\underline{A}^T\underline{D}^{-1}\underline{1}) - \underline{A}^T\underline{D}^{-2}\underline{1}\}$$

$$-(P^2-P)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) + (\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}(\underline{A}^T\underline{D}^{-1}\underline{1})(\underline{1}^T\underline{D}^{-1}\underline{1}) \{(\underline{1}^T\underline{D}^{-1}\underline{1})^{-2} - (\underline{1}^T\underline{D}^{-2}\underline{1})\}\}/(N-1)K$$

$$+\lambda_1\{(P-1)(\underline{1}^T\underline{D1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) - 2(P^2-P)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1})$$

$$-2\{(\underline{1}^T\underline{D1})^2 - (\underline{1}^T\underline{D}^2\underline{1})\}(\underline{1}^T\underline{D}^2\underline{1})\}(\underline{1}^T\underline{D}^{-1}\underline{1})(\underline{A}^T\underline{D}^{-1}\underline{1})/K$$

$$-2(\underline{1}^T\underline{D1})^2(\underline{A}^T\underline{D}^{-1}\underline{1})(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}\{\underline{1}^T\underline{D}^{-1}\underline{1})^{-2} - (\underline{1}^T\underline{D}^{-2}\underline{1})\}/K$$

$$+4(P^2-P)(\underline{1}^T\underline{D1})(\underline{A}^T\underline{D}^{-1}\underline{1})/K\}/(N-1)K .$$

In the above, if we ignore the terms of 1/(N-1) order, we have

$$\underline{\hat{B}}^{**} \approx (1-\lambda_1)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}), \tag{2.12}$$

where

$$\lambda_1 = cP^2/\{(\underline{1}^T\underline{D}\underline{1})(\underline{1}^T\underline{D}^{-1}\underline{1}) - P^2\}.$$

We note that if $\lambda_1 = 0$, then

$$\underline{\hat{B}}^{**} \approx (\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) \tag{2.13}$$

and if we use this as an estimator of $\underline{B}$,

$$\hat{\mu} = \overline{Y_1} + (\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}\underline{A})\overline{\underline{Y_2}}$$

$$= \underline{1}^T\overline{\underline{X}}/P + (\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}\underline{A})\underline{A}^T\overline{\underline{X}}.$$

Since $\underline{A}\underline{A}^T = \underline{I} - \underline{1}\underline{1}^T/P$, we get

$$\hat{\mu} = (\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}\overline{\underline{X}}) = \hat{\mu}_G.$$

That is, Graybill-Deal estimator use (2.13) as an estimator of $\underline{B}$. For our convenience, we define

$$\lambda = cP^2/\{(\underline{1}^T\underline{D}\underline{1})(\underline{1}^T\underline{D}^{-1}\underline{1})\},$$

where $0 \le c \le 1$, as a shrinkage parameter. From this, we may construct a new estimator of μ ,

$$\hat{\mu}_{NEW} = (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}\overline{\underline{X}}) + \lambda\underline{1}^T\overline{\underline{X}}/P, \tag{2.14}$$

to achieve our purpose, improvements on $\hat{\mu}_G$. Note that we do not assume special conditions in (2.14), though Baranchik assumes $P \geq 4$ and $N \geq P+1$.

As another shrinkage parameter, we can choose

$$\lambda^* = c(\Pi S_i^{\,2})^{2/P}/(\sum S_i^{\,2}/P)^2,$$

which is mimicked the test statistics for equality of several variances. Hence as an estimator of μ, we may suggest

$$\hat{\mu}_{NEW1} \equiv (1-\lambda^*)\hat{\mu}_G + \lambda^* \overline{\mu}\ .$$

3. Variance of new unbiased estimator

In this section, we investigate the unbiasedness and variance of new estimator.

Since $S_i^{\,2}$'s are distributed independently of the $\overline{X_i}$'s and the conditional expectation of $\hat{\mu}_G$ given $S_i^{\,2}$'s is μ, unbiasedness of $\hat{\mu}_{NEW}$ is directly derived as follows

$$\begin{aligned} E\{\hat{\mu}_{NEW}\} &= E_{S_1^2 \cdots S_P^2} E_{\overline{X_1}, \cdots \overline{X_P} \mid S_1^2 \cdots S_P^2}\{ (1-\lambda)\hat{\mu}_G + \lambda\overline{\mu}\} \\ &= E_{S_1^2 \cdots S_P^2}\{ (1-\lambda)\mu + \lambda\mu\} \\ &= \mu. \end{aligned}$$

From (2.8),

$$Var(\hat{\mu}_{NEW}) = Var(\hat{\mu}^*) + \frac{1}{N} E\{ (\underline{\hat{B}}^{**} - \underline{B})^T \underline{V}_{\underline{Y_2 Y_2}} (\underline{\hat{B}}^{**} - \underline{B})\}. \tag{3.1}$$

The second term of (3.1) becomes

$$\begin{aligned} & E\{ \{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{V}^{-1}\underline{1})\}^T \\ & \qquad \underline{A}^T\underline{V}\underline{A}\{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{D}^{-1}\underline{1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{A}^T\underline{V}^{-1}\underline{1})\} \} \\ & = E\{ \{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{V}^{-1})\} \\ & \qquad \underline{AA}^T\underline{V}\underline{AA}^T\{ \{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{D}^{-1}\underline{1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{V}^{-1}\underline{1})\} \}. \end{aligned} \tag{3.2}$$

Since $\underline{AA}^T = \underline{I} - \underline{11}^T/P$, it becomes

$$\begin{aligned} & E\{ \{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{V}^{-1}) + \lambda\underline{1}^T/P\} \\ & \qquad \underline{V}\{ (1-\lambda)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{D}^{-1}\underline{1}) - (\underline{1}^T\underline{V}^{-1}\underline{1})^{-1}(\underline{V}^{-1}\underline{1})\} + \lambda\underline{1}/P \}. \end{aligned}$$

Hence we obtain

$$Var(\hat{\mu}_{NEW}) = \frac{1}{N} E\{ (\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}(\underline{1}^T\underline{D}^{-1}\underline{V}\underline{D}^{-1}\underline{1}\}$$

$$+\frac{1}{N}E\Big\{2(1-\lambda)\lambda\frac{1}{P}(\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}(\underline{1}^T\underline{D}^{-1}\underline{V}\underline{1})$$

$$-(2\lambda-\lambda^2)(\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}(\underline{1}^T\underline{D}^{-1}\underline{V}\underline{D}^{-1}\underline{1})+\frac{\lambda^2}{P^2}(\underline{1}^T\underline{V}\underline{1})\Big\}$$

$$= Var(\hat{\mu}_G)+\frac{1}{N}E\Big\{\frac{2\lambda}{P}\{((\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}((\underline{1}^T\underline{D}^{-1}\underline{V}\underline{1})$$

$$-P((\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}(\underline{1}^T\underline{D}^{-1}\underline{V}\underline{D}^{-1}\underline{1})\}+\lambda^2\{P^{-2}(\underline{1}^T\underline{V}\underline{1})$$

$$+(\underline{1}^T\underline{D}^{-1}\underline{1})^{-2}((\underline{1}^T\underline{D}^{-1}\underline{V}\underline{D}^{-1}\underline{1})-\frac{2}{P}((\underline{1}^T\underline{D}^{-1}\underline{1})^{-1}((\underline{1}^T\underline{D}^{-1}\underline{V}\underline{1})\}\Big\} \quad (3.3)$$

We note that if the second term of (3.3) is a negative quantity, then $\hat{\mu}_{NEW}$ dominates $\hat{\mu}_G$ in variance. In case of equal variances, we can show that the second term of (3.3) is negative. That is, when all σ_i^2's are 1,

$$E\Big\{2\lambda P^{-1}\Big\{1-P\sum_{i=1}^{P}\frac{1}{S_i^4}\Big/(\sum_{i=1}^{P}\frac{1}{S_i^2})^2\Big\}$$

$$+\lambda^2P^{-2}\Big\{P+P^2\sum_{i=1}^{P}\frac{1}{S_i^4}\Big/(\sum_{i=1}^{P}\frac{1}{S_i^2})^2-2P\Big\}\Big\}$$

$$=E\{2P^{-1}\lambda-2\lambda A+\lambda^2A-P^{-1}\lambda^2\}$$

$$=E\{(\lambda^2-2\lambda)(A-P^{-1})\}$$

where $A\equiv\sum_{i=1}^{P}\frac{1}{S_i^4}\Big/(\sum_{i=1}^{P}\frac{1}{S_i^2})^2$. Since $0<\lambda\le 1$, and $A\ge P^{-1}$, the last equation becomes negative. Hence new estimator dominates Graybill-Deal estimator when all variances are equal.

For P=2 and N=3, some analytic results are obtained by Lee(1993). But for P>2 it is very difficult to utilize the analytic techniques to evaluate (3.3) exactly. In the next section, we introduce the simulation results about efficiency of new estimator and that of Graybill-Deal.

4. Monte Carlo simulation study

In this section, we carry out Monte Carlo simulation and investigate efficiency of new estimator and Graybill-Deal estimator.

The computer programs used for simulation were written in Fortran 77 and were run on the CIBER962-31 computer at Kangwon National University. Uniform random numbers used in these simulations were generated by using the IMSL.

For the sake of estimating variances of new estimator, (3.3) by Monte Carlo simulations, from the facts that each S_i^2/σ_i^2, $i=1, \cdots, P$, have chi-square distribution with (N-1) degrees of freedom, we first generate chi-square variates and obtain random variates S_i^2. We confirm our attention to odd sample numbers. Then chi-square variates with (N-1) degrees of freedom are generated as like $Y \equiv -2\mathrm{Ln}(\Pi U_i)$, using the (N-1)/2 uniform random variates $U_i s$. The same experiments were conducted 10,000 times. We define $\widehat{Var}(\hat{\mu}_G)$ and $\widehat{Var}(\hat{\mu}_{NEW})$ to be variances obtained by simulations. We define

$$\widehat{eff}(\hat{\mu}_G) \equiv Var(\hat{\mu}^*)/\widehat{Var}(\hat{\mu}_G)$$

$$\widehat{eff}(\hat{\mu}_{NEW}) \equiv Var(\hat{\mu}^*)/\widehat{Var}(\hat{\mu}_{NEW})$$

$$\widehat{eff}(\hat{\mu}_{NEW1}) \equiv Var(\hat{\mu}^*)/\widehat{Var}(\hat{\mu}_{NEW1}) \ .$$

When P=4-12, we consider the several cases for population's variances and show the results for the comparisons of efficiencies in figure 1-10. In figures, NEW, NEW1 and GRAYBILL denote the efficiency of their estimators. We took c=(N-2)/(N-1) in figures 1-6 and c=2/(N-1) in figures 7-14.

Figure 1. Comparison of efficiency when $\sigma_i^2=1$, $i=1, \cdots, 4$.

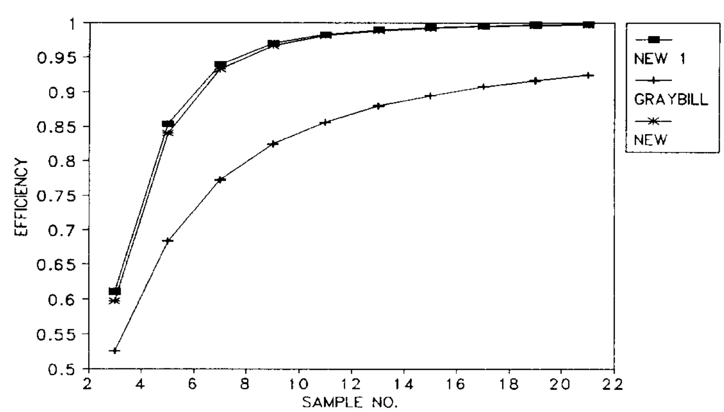


Figure 2. Comparison of efficiency when $\sigma_i^2=1,\ i=1,\cdots,10.$

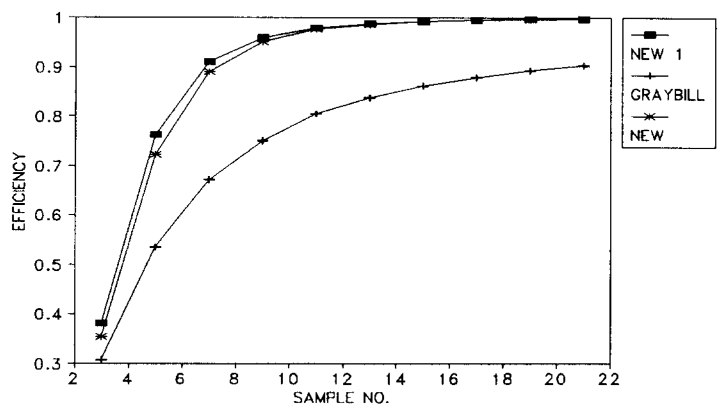


Figure 3. Comparison of efficiency when $\sigma_i^2=i,\ i=1,\cdots,4.$

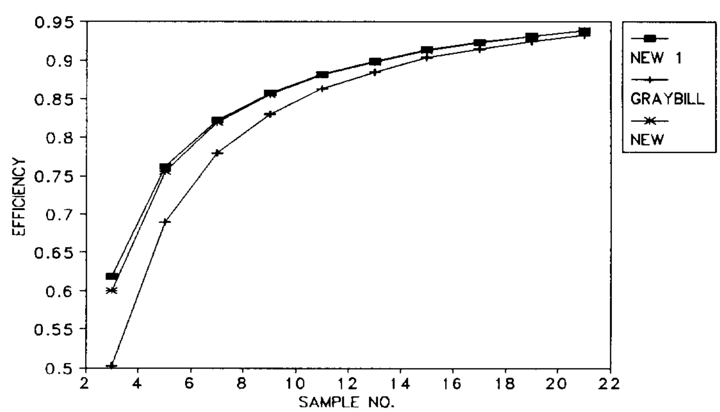


Figure 4. Comparison of efficiency when $\sigma_i^2 = i$, $i = 1, \cdots, 6$.

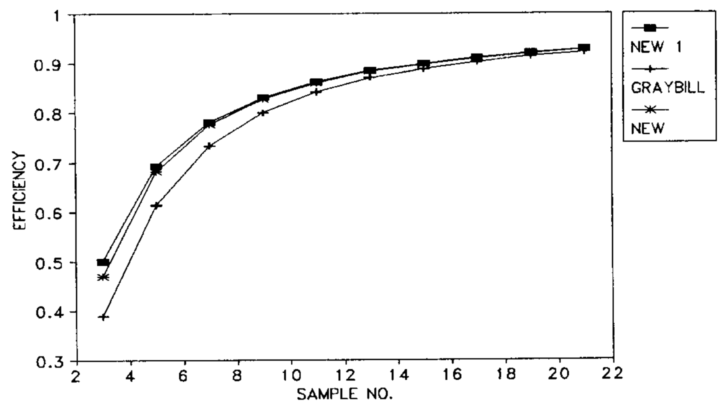


Figure 5. Comparison of efficiency when $\sigma_i^2 = 1$, $i = 1, \cdots, 3$, $\sigma_4^2 = 7$.

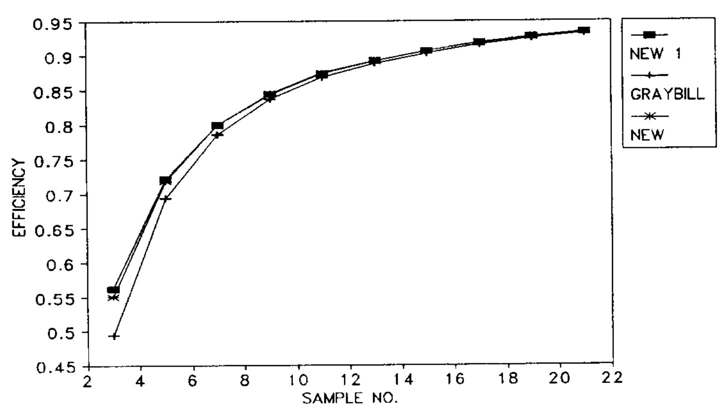


Figure 6. Comparison of efficiency when $\sigma_1^2=1$, $\sigma_2^2=1$, $\sigma_3^2=5$, $\sigma_4^2=5$.

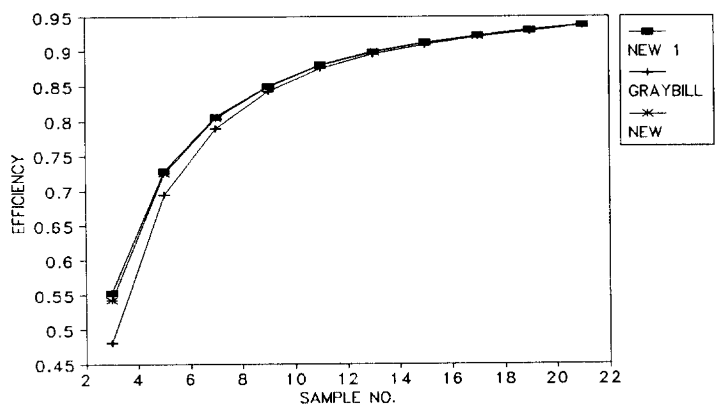


Figure 7. Comparison of efficiency when $\sigma_1^2=1$, $\sigma_2^2=8$, $\sigma_3^2=8$, $\sigma_4^2=8$.

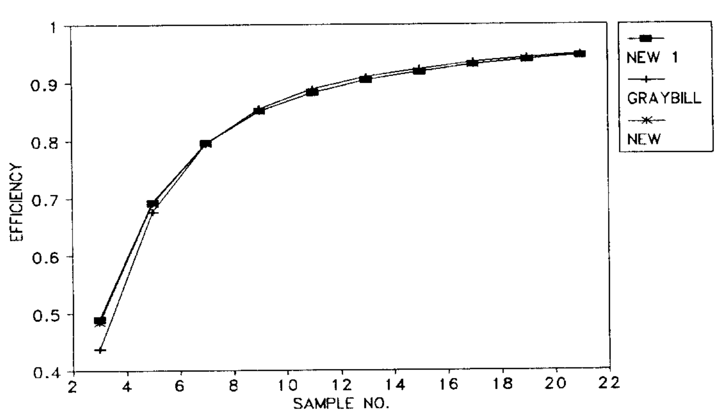


Figure 8. Comparison of efficiency when $\sigma_i^2=1,\ i=1,\cdots,7,\ \sigma_8^2=10.$

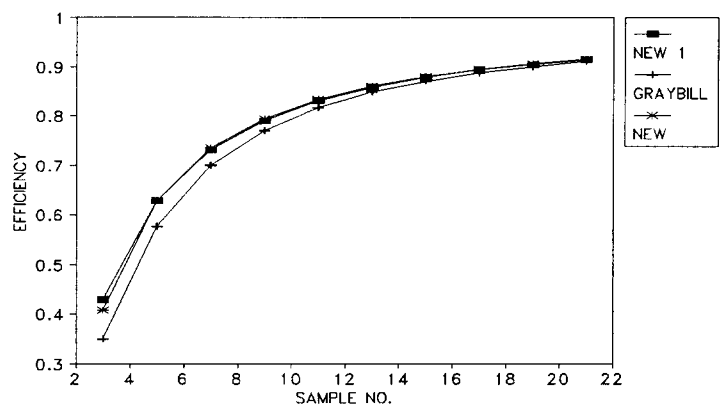


Figure 9. Comparison of efficiency when $\sigma_1^2=1,\ \sigma_i^2=8,\ i=2,\cdots,\ 8.$

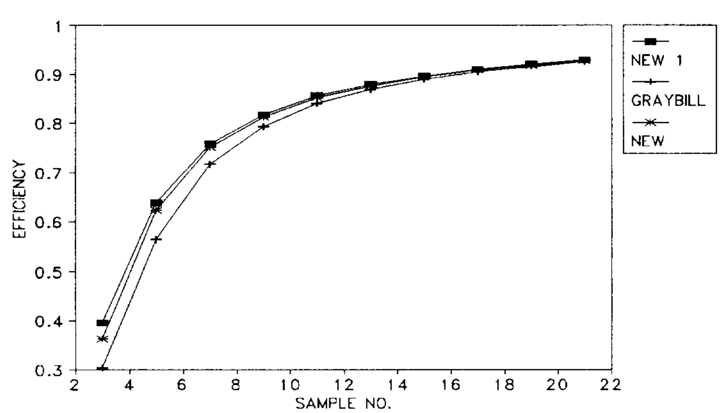
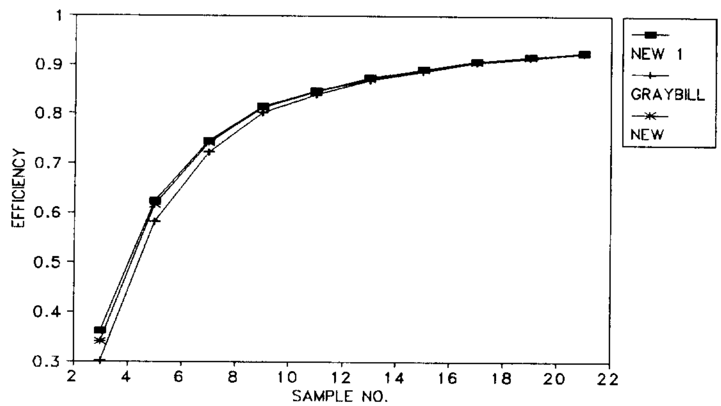


Figure 10. Comparison of efficiency when
$\sigma_1{}^2=1,\ \sigma_2{}^2=1,\ \sigma_3{}^2=4,\ \sigma_4{}^2=4,\ .\sigma_5{}^2=8,\ \sigma_6{}^2=8,\ \sigma_7{}^2=12,\ \sigma_8{}^2=12,$



From the results we can conclude that a class of new estimators presented in this paper has better performance than Graybill-Deal estimator for all sample sizes except the case when some of the $\sigma_i{}^2$'s are very small compared with the other. We note $\hat{\mu}_{NEW1}$ has better performance than $\hat{\mu}_{NEW}$.

5. Conclusions

When $P \geq 3$, it is very difficult to treat new estimator analytically. From Monte-Carlo simulation, although it may be pointed out that the gains in efficiency tends to decrease when sample size increase, it also turns out that the efficiency of new estimator is better than that of Graybill-Deal estimator when the variances of the population are almost the same or when we have small sample size.

REFERENCES

[1] Baranchik, A. J. (1973). Inadmissibility of maximum likelihood estimator in some multiple regression problems with three or more independent variables, *The annals of Statistics*, Vol. 1, 312-321.

[2] Bhattacharya, C. G. (1978). Yates type estimators of a common mean, *Annals of The Institue of Statistical Mathematics*, Vol. 30, Part A, 407-414.

[3] Bickel, P. J. and Doksum, K. A.(1977). *Mathematical Statistics-Basic Ideas and Selected Topic-*, Holden-Day, Inc., San Francisco.

[4] Brown, L. D. and Cohen, A.(1974). Point and confidence estimation of a common mean and recovery of interblock information, *The annals of Statistics*, Vol. 2, 963-976.

[5] Cohen, A. and Sackrowitz, H. B.(1974). On estimating the common mean of two normal distributions, *The annals of Statistics,* Vol. 2, 1274-1282.

[6] Graybill, F. A. and Deal, R. B.(1959). Combining unbiased estimators, *Biometrics*, Vol. 15, 543-550.

[7] Lee, S. S.(1992). New unbiased estimator of a common mean of two normal populations, *The Journal of Science and Technology*, Kangwon National University, Vol. 31, 162-166.

[8] Metha, J. S. and Gurland, J.(1969). Combinations of unbiased estimation of the mean which consider inequality of unknown variances, *Journal of American Statistical Association*, Vol. 64, 1042-1055.

[9]Norwood, Jr. T. E. and Hinkelmann, K.(1977). Estimating the common mean of several normal populations, *The annals of Statistics,* Vol. 5, 1047-1050.

[10]Rao, Jr. N. K.(1980). Estimating the common mean of possibly different normal populations: A simulation study, *Journal of American Statistical Association,* Vol. 75, 447-453.

[11]Shina, B. K. and Mouqadem, O.(1982). Estimation of the common mean of two univariate normal populationa, *Communication in Statistics, Part A-Theory and Method*, Vol. 11, 1603-1614.

[12]Shinozaki, N.(1978). A note on estimating the common mean of K normal distribution and the Stein problem, *Communication in Statistics, Part A-Theory and Method,* Vol. 7, 1421-1432.

[13]Zacks, S.(1966). Unbiased estimation of the common mean of two normal distributions based on small samples of equal size, *Journal of American Statistical Association* Vol. 61, 467-476.