SOME MULTIPLE DECISION PROBLEMS
IN ANALYSIS OF VARIANCE*

by

Shanti S. Gupta, Purdue University and
Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #458

July 1976

*This research was supported by the Office of Naval Research Contracts
NOOO14-67-A-0226-0014 and N00014-75-C-0455 at Purdue University.
Reproduction in whole or in part is permitted for any purpose of the
United States Government.
SOME MULTIPLE DECISION PROBLEMS
IN ANALYSIS OF VARIANCE*

by

Shanti S. Gupta, Purdue University and
Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

1. Introduction

In most practical situations to which the analysis of variance tests
are applied, they do not supply the information that the experimenter aims
at. If, for example, in one-way ANOVA the hypothesis is rejected in
actual application of the F-test, the resulting conclusion that the true
means $\theta_1, \theta_2, \ldots, \theta_k$ are not all equal, would by itself usually be insufficient
to satisfy the experimenter. In fact his problems would begin at this stage.
The experimenter may desire to select the "best" population or a subset of
the "good" populations; he may like to rank the populations in order of
"goodness" or he may like to draw some other inferences about the parameters
of interest.

The extensive literature on selection and ranking procedures depends
heavily on the use of independence between populations (block, treatments,
etc.) in the analysis of variance. In practical applications, it is
desirable to drop this assumption of independence and consider cases more
general than the normal.

Our interest is to derive a method to construct locally best (in some
sense) selection procedures to select a nonempty subset of the $k$ populations

*This research was supported by the Office of Naval Research Contracts
N00014-67-A-0226-0014 and N00014-75-C-0455 at Purdue University. Reproduction
in whole or in part is permitted for any purpose of the United States
Government.
containing the best population as ranked in terms of \( \theta_i \)'s (defined below) which control the size of the selected subset and maximizing the probability of selecting the best. We also consider the usual selection procedures in one way ANOVA based on the generalized least square estimates and apply the method to two way layout case. Some examples are discussed and some results on comparisons with other procedures are also considered.

2. Locally Best Selection Procedures

Let \( \pi_1, \pi_2, \ldots, \pi_k \) represent \( k (\geq 2) \) populations and let \( X_{i1}, \ldots, X_{in_i} \) be \( n_i \) independent random observations from \( \pi_i \). The selection procedures will depend upon the observations through \( T_{ij} \) which are defined as follows.

Let \( T_{ij} = T(X_{i1}, \ldots, X_{in_i}; X_{j1}, \ldots, X_{jn_j}) \) be based on the \( n_i \) and \( n_j \) observations from \( \pi_i \) and \( \pi_j \) \( (i, j = 1, 2, \ldots, k) \), respectively. In a given problem the function \( T \) is so chosen as to indicate the differences between the populations in a reasonable way. For example, if the observations drawn from \( \pi_i \) are normally distributed with unknown mean \( \theta_i \), \( (1 \leq i \leq k) \), and known variance \( \sigma^2 \), a choice of \( T_{ij} \) might be \( X_i - X_j \), where

\[
X_i = \frac{1}{n_i} \sum_{l=1}^{n_i} X_{il} \quad \text{and} \quad X_j = \frac{1}{n_j} \sum_{l=1}^{n_j} X_{jl}.
\]

Now we assume that \( T_{ij} \) has a joint probability density function \( g_{\tau_{ij}} (\cdot) \) depending on the parameter \( \tau_{ij} \) and assume that \( \tau_{ij} \)'s are known. Usually \( T_{ij} \)'s are chosen to obtain both sufficient and maximal invariant statistics for \( \tau_{ij} \)'s. Let \( \tau_i = \min_{j \neq i} \tau_{ij} \). Returning to the above normal means problem, we find that \( \tau_{ij} = \theta_i - \theta_j \) and \( \tau_i = \theta_i - \theta[k] \) for \( \theta_i < \theta[k] \) and \( \tau_i = \theta_i - \theta[k-1] \) for \( \theta_i = \theta[k] \), where

\[ \theta[1] \leq \ldots \leq \theta[k]. \]

A population is said to be best if \( \tau_i = \max_{1 \leq j \leq k} \tau_j \). For the above normal means example, \( \pi_i \) is best if \( \theta_i = \theta[k] \) and in this case \( \tau_{ii} = 0 \) and \( \tau_i = \theta[k] - \theta[k-1] \).
Let \( \xi_1 = (\tau_{ij} | 1 \leq j \leq k, j \neq i) \), \( \xi_1^i = (\tau_{ij} | \tau_{ii} = \tau_{ij}, 1 \leq i \leq k, i \neq j) \) and \( z_i = (z_{ij} | 1 \leq j \leq k, j \neq i) \) are all \( k-1 \) dimensional vectors. Assume that the joint density of \( T_{ij}, j = 1,2,\ldots,k, j \neq i \), is \( f_{\xi_1}(z_1) \), \( 1 \leq i \leq k \), (with respect to some \( \sigma \)-finite measure \( \mu \)). Let \( \delta_i \) be the probability of selecting \( \pi_i \),

\[
S_r = \{ \delta: \delta = (\delta_1, \ldots, \delta_k), \sum_{i=1}^{k} \int_{\xi_1} f_{\xi_1}(z_1) d\mu(z_1) \leq r \}
\]

and

\[
S_r' = \{ \delta: \delta = (\delta_1, \ldots, \delta_k), \sum_{i=1}^{k} \int_{\xi_1} f_{\xi_1}(z_1) d\mu(z_1) = r \}.
\]

**Theorem:** Let \( \delta^0 = (\delta^0_1, \ldots, \delta^0_k) \in S_r' \) be defined by

\[
\delta^0_i(z_1) = \begin{cases} 
1 & \text{if } \min_{1 \leq j < k} \frac{a}{\partial \tau_{ij}} f_{\xi_1}(z_1) |_{\xi_1} > cf_{\xi_1}(z_1), \\
\lambda_i & < \\
0 & \text{if } \min_{1 \leq j < k} \frac{a}{\partial \tau_{ij}} f_{\xi_1}(z_1) |_{\xi_1} \leq cf_{\xi_1}(z_1)
\end{cases}
\]

Then \( \delta^0 \) maximizes \( \sum_{i=1}^{k} \int_{\xi_1} f_{\xi_1}(z_1) d\mu(z_1) \) among all rules \( \delta \in S_r \). \( \delta^0 \) is called a locally best procedure in this sense.

**Proof.** For any \( \delta \in S_r \),

\[
\sum_{i=1}^{k} \int_{\xi_1} f_{\xi_1}(z_1) \min_{j \neq i} \frac{a}{\partial \tau_{ij}} f_{\xi_1}(z_1) |_{\xi_1} d\mu(z_1)
\]

\[
- \sum_{i=1}^{k} \int_{\xi_1} f^0_{\xi_1}(z_1) \min_{j \neq i} \frac{a}{\partial \tau_{ij}} f_{\xi_1}(z_1) |_{\xi_1} d\mu(z_1)
\]

\[
= \sum_{i=1}^{k} \int f([\delta_1(z_1) - \delta^0_1(z_1)] \min_{j \neq i} \frac{a}{\partial \tau_{ij}} f_{\xi_1}(z_1) |_{\xi_1} - c f_{\xi_1}(z_1)) d\mu(z_1)
\]

\[
+ c \sum_{i=1}^{k} \int [f(z_1) - \delta^0_1(z_1)] f_{\xi_1}(z_1) d\mu(z_1) \leq 0.
\]

This proof is complete.
Example: Let $g_{\xi}(x) = \prod_{j=1}^{k} g_{\theta_i}(\bar{x}_i)$, where $g_{\theta_i}(\bar{x}_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (\bar{x}_i - \theta_i)^2}$. Let $	au_{ij} = \theta_i - \theta_j$, $1 \leq j \leq k$, $j \neq i$, $	au_{ii} = 0$, and $Z_{ij} = \bar{x}_i - \bar{x}_j$, $j \neq i$. We know that a maximal invariant under a group $G$ is $T_i = (\bar{x}_i - \bar{x}_1, \ldots, \bar{x}_i - \bar{x}_k)$ where $G$ is the group of transformations

$$g_{Z_i} = (z_{i1} + c, \ldots, z_{ik} + c), \quad -\infty < c < \infty,$$

which, in the parameter space, induces the transformations $g_{\tau_{ij}} = \tau_{ij} + c$.

Since $\Sigma(k-1)x(k-1) = \frac{1}{n} (\Sigma \Sigma^{-1})$ is the covariance matrix of $Z_{ij}$'s. We know that $\Sigma$ is positive definite, and

$$\Sigma^{-1} = \frac{n}{k} \begin{pmatrix} k-1 & & & \\ & k-1 & & \\ & & \ddots & \\ & & & k-1 \end{pmatrix}.$$  

Hence

$$f_{\xi_i}(z_i) = (2\pi)^{-\frac{k-1}{2}} |\Sigma|^{-\frac{1}{2}} \exp{(-\frac{n}{k} \frac{1}{2} \left( (z_i - \tau_i) \Sigma^{-1} (z_i - \tau_i) \right)^2})$$

$$-\ldots - (z_i_1 - \tau_i_1)(z_i_k - \tau_i_k) + \ldots + (-z_i_1 - \tau_i_1)(z_i_k - \tau_i_k)$$

$$-\ldots - (z_i(k-1) - \tau_i(k-1))(z_i(k-1) - \tau_i(k-1)) + (k-1)(z_i(k-1) - \tau_i(k-1))^2).$$

Thus

$$\frac{\partial}{\partial \tau_{ij}} f_{\xi_i}(z_i) \bigg|_{\xi_i = \xi_i} = (2\pi)^{-\frac{k-1}{2}} |\Sigma|^{-\frac{1}{2}} \exp{(-\frac{n}{k} \frac{1}{2} \left( (k-1)z_i^2 + z_i z_i^2 - \ldots - z_i z_i^2 \right)}} + \ldots + (-z_i z_i^2 + \ldots - z_i(k-1)z_i^2$$

$$+ (k-1)z_i^2 \right) \cdot (-2z_i - \ldots - 2z_i(j-1) + 2(k-1)z_i j - 2z_i(j+1)^2$$

$$\ldots - 2z_i k).$$

Hence
\[
\delta_i^0(z_i) = \begin{cases} 
1 & \text{if } \min_{j \neq 1} \left[ -\sum_{j=1, j \neq i, j}^{k} z_{ij} + (k-1)z_{ij} \right] \geq c, \\
0 & \text{otherwise}
\end{cases}
\]

or

\[
\delta_i^0(z_i) = \begin{cases} 
1 & \text{if } \max_{1 \leq j \leq k} \bar{x}_j \leq \frac{1}{k} \sum_{j=1}^{k} \bar{x}_j - c, \\
0 & \text{otherwise}
\end{cases}
\]

3. Usual Approach to Selection Problems in One Way Layout

Let \( \pi_1, \pi_2, \ldots, \pi_k \) be \( k \) populations. Let \( X_{i1}, \ldots, X_{in_i} \) denote \( n_i \) independent observations from the \( i \)th population \( \pi_i \). Let the joint density of \( X_{11}, \ldots, X_{1n_1}; X_{21}, \ldots, X_{2n_2}; \ldots; X_{k1}, \ldots, X_{kn_k} \) be of the following form:

\[
(3.1) \quad c_k |\Lambda|^{-1} g((x_\mathbf{-d})', \Lambda^{-1}(x_\mathbf{-d}))
\]

where \( x' = (x_{11}, \ldots, x_{1n_1}; \ldots; x_{k1}, \ldots, x_{kn_k}) \), \( n' = (\theta_1, \ldots, \theta_1; \ldots; \theta_k, \ldots, \theta_k) \)

and \( \Lambda \) is a known positive definite matrix and \( c_k \) is determined such that (3.1) is a density.

Let \( \Omega = \{ \Theta : \Theta' = (\theta_1, \ldots, \theta_k) \} \) and also, let

\[
A_{nk} = \begin{pmatrix}
1_{n_1} & 0 \\
1_{n_2} & \ddots \\
0 & \ddots & 1_{n_k}
\end{pmatrix},
\]
where \( \xi_n^i = (1, \ldots, 1) \) with \( n_i \) components and \( \sum_{i=1}^{k} n_i = N > k \).

We consider the analysis of variance problem in a one way layout;

let

\[
x_{ij} = \theta_i + e_{ij}, \quad j=1, \ldots, n_i; \quad i=1, \ldots, k
\]

which is in the form of the general linear model

\[
x = A\theta + e,
\]

where \( e' = (e_{11}, \ldots, e_{1n_1}; \ldots; e_{k1}, \ldots, e_{kn_k}) \) with \( \text{var}(e) = \Lambda \).

We know that the generalized least square estimator of \( \theta \) is

\[
\hat{\theta} = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}x = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = y.
\]

Since \( \hat{\theta} = Bx \), \( B = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \), the joint density of \( Y' = (Y_1, \ldots, Y_k) \) is of the form

\[
b_k|\Lambda|^{-\frac{1}{2}} h((Y-\theta)'\Lambda^{-1}(Y-\theta))
\]

where \( \Lambda_1 = BAB' = (\sigma_{ij}) \).

The ordered \( \theta_i \)'s are denoted by \( \theta[1] \leq \ldots \leq \theta[k] \). It is assumed that there is no prior knowledge of the correct pairing of the ordered and the unordered \( \theta_i \)'s. Let us denote by \( \pi(i) \) the population (unknown) associated with \( \theta[i_1], i = 1, 2, \ldots, k \). Our goal is to select a non-empty subset of the \( k \) populations so as to include the population associated with \( \theta[k] \). Defining any such selection as a correct selection, we wish to define a procedure \( R \) so that \( P(\text{CS}|R) \), the probability of a correct selection, is at least a preassigned number \( P^*(\frac{1}{k} < P^* < 1) \). We will refer to this requirement as the \( P^* \)-condition. We propose the following rule \( R \) based on \( Y_i, 1 \leq i \leq k \).
R: Retain \( \pi_i \) in the selected subset if and only if

\[
Y_i \geq \max_{1 \leq j \leq k} \left( Y_j - c \sqrt{\sigma_{i1} + \sigma_{jj} - 2\sigma_{ij}} \right),
\]

where \( c = c(k, p^*; n_i, \sigma_{ij}, 1 \leq i, j \leq k) > 0 \) is chosen so as to satisfy the \( p^* \)-condition.

Let \( Y(i) \) and \( \sigma(i)(i) \) denote the observation and the variance associated with the population \( \pi(i) \) with mean \( \theta[i] \), \( i = 1, 2, \ldots, k \). Of course, both \( Y(i) \) and \( \sigma(i)(i) \) are unknown as in \( \sigma(i)(j) \), the covariance of \( Y(i) \) and \( Y(j) \). Thus

\[
(3.3) \quad P(CS|R) = P\{Y(k) \geq \max_{1 \leq j \leq k} \left( Y(j) - c \sqrt{\sigma(k)(k) + \sigma(j)(j) - 2\sigma(k)(j)} \right) \}
= P\{Z_{jk} \leq c(\theta_k - \theta_j)(\sigma(k)(k) + \sigma(j)(j) - 2\sigma(k)(j))^{\frac{1}{2}}, 1 \leq j \leq k-1\},
\]

where for \( \lambda, 1 \leq \lambda \leq k \), we define

\[
(3.4) \quad Z_{\rho\lambda} = (Y(r) - Y(\rho)) - \theta[r] + \theta[\rho] \left( \sigma(r)(r) + \sigma(\rho)(\rho) - 2\sigma(r)(\rho) \right)^{-\frac{1}{2}},
\]

for \( r = 1, 2, \ldots, k \), \( r \neq \lambda \).

Let \( Z_{\rho\lambda} = Y A_{\rho\lambda} \), where \( Z_{\rho\lambda} = (Z_{r\rho} : r = 1, 2, \ldots, k, r \neq \rho) \) and

\[
Y = (Y(1) - \theta[1], \ldots, Y(k) - \theta[k]).
\]

The matrix \( A_{\rho\lambda} \) with \( k \) rows and \( (k-1) \) columns is defined as follows:

Let \( \alpha_{r\rho} = (\sigma(r)(r) + \sigma(\rho)(\rho) - 2\sigma(r)(\rho))^{-\frac{1}{2}}, 1 \leq r, \rho \leq k, r \neq \rho; \)

\[
A_{\rho\lambda} = \begin{pmatrix}
\alpha_{1\rho} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \alpha_{2\rho} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & \alpha_{\rho-1\rho} & 0 & \cdots & 0 \\
-\alpha_{1\rho} & -\alpha_{2\rho} & \cdots & -\alpha_{\rho-1\rho} & -\alpha_\rho & \cdots & -\alpha_k \\
0 & 0 & \cdots & 0 & \alpha_{\rho+1\rho} & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & \alpha_{k\rho}
\end{pmatrix}
\]
and \( \Sigma_1 = (\sigma_{ij}(j)) \).

Since \( A_k^t \Sigma_1 A_k = (\rho_{ij}(k)) \), \( i, j = 1, 2, \ldots, k \); \( i, j \nmid k \), for \( 1 \leq \ell \leq k \) and \( A_k \) is of rank \( k-1 \), hence the joint density of \( Z_k \) is

\[
\mathcal{L}(Z_k) = a_k |A_k^t \Sigma_1 A_k|^{-\frac{1}{2}} f(z_k^t (A_k^t \Sigma_1 A_k)^{-1} z_k).
\]

(3.6)

For any given association between \((\sigma_{ij}, i, j=1,2,\ldots,k)\) and \((\sigma_{ij}(j); i, j=1,2,\ldots,k)\) we can see from (3.3) that the infimum of \( P(CS|R) \) is attained when \( \theta[1] = \ldots = \theta[k] \).

Hence,

\[
\inf_{\theta \in \Omega} P(CS|R) = \min_{1 \leq \ell \leq k} P(Z_{r\ell} \leq c, r=1,2,\ldots,k, r\nmid \ell; \{\rho_{ij}(\ell)\}).
\]

(3.7)

For \( \ell < k \), let \( \kappa_{ij}^\ell \) be such that

(i) \( \kappa_{ij}^\ell \geq \kappa_{ij}^k \), \( i, j \nmid k \), \( i, j \nmid k \);

(3.8)

(ii) \( \kappa_{ij}^\ell \geq \kappa_{ij}^k \), \( j=1,2,\ldots,k \); \( j \nmid \ell \), \( k \),

and for any \( \ell \), \( 1 \leq \ell \leq k \), there exists an \( m \), \( 1 \leq m \leq k \), such that for any \( i, j, i, j=1,2,\ldots,k \), \( i, j \nmid k \), \( i \nmid j \),

\[
\kappa_{ij}^\ell = \rho_r(m)
\]

for some \( r, s, r \nmid s, r, s \nmid m, r, s=1,2,\ldots,k \).

Lemma [3]. If \( X = (X_1, \ldots, X_p) \) has density \( |\Sigma|^{-\frac{1}{2}} f(x^t \Sigma^{-1} x) \), then for any two positive definite (symmetric) pxp matrices \( \Gamma_1 = (r_{ij}) \) and \( \Gamma_2 = (\sigma_{ij}) \) such that \( r_{ii} = \sigma_{ii}, 1 \leq i \leq p \) and \( r_{ij} \geq \sigma_{ij}, 1 \leq i < j \leq p \),

\[
P_{\Gamma_1}(X_1 < \ell_1, \ldots, X_p < \ell_p) \geq P_{\Gamma_2}(X_1 < \ell_1, \ldots, X_p < \ell_p)
\]

for any real numbers \( \ell_1, \ldots, \ell_p \).
By (3.8) and Lemma, we have

\begin{equation}
(3.9) \quad \inf_{\Theta \in \Omega} P_{\Theta}(CS|R) \\
= \min_{1 \leq \ell \leq k} P(Z_{r \ell} \leq c, r=1,2,\ldots,k; r \neq \ell; \{\kappa_{ij}^k\}) \\
= P(Z_{rk} \leq c, r=1,2,\ldots,k-1; \{\kappa_{ij}^k\}) \\
= \int_{-\infty}^{c} \cdots \int_{-\infty}^{c} a_k |(\kappa_{ij}^k)|^{-\frac{\lambda}{2}} f(z_k'((\kappa_{ij}^k)^{-1}z_k) dz_1 \cdots dz_{k-1}.k.
\end{equation}

Discussion of Condition (3.8)

For computational convenience, we assume that \(A = \text{diag}(\lambda_{11}, \ldots, \lambda_{11}, \ldots, \lambda_{kk}), \lambda_{ii} > 0, i=1,2,\ldots,k\). If also the components \(X_{11}, \ldots, X_{1n_1}; \ldots; X_{k1}, \ldots, X_{kn_k}\) are independent then the joint distribution \(g\) as in (3.1) is multivariate normal (see Kelker [9, p. 18]). Then

\[ A' \Lambda^{-1} A = \text{diag} \left( \frac{n_1}{\lambda_{11}}, \frac{n_2}{\lambda_{22}}, \ldots, \frac{n_k}{\lambda_{kk}} \right) \text{ and} \]

\[ B = (A' \Lambda^{-1} A)^{-1} A' \Lambda^{-1} = \begin{pmatrix}
\frac{1}{n_1} & 0 & \cdots & 0 \\
0 & \frac{1}{n_2} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & \frac{1}{n_k} \\
0 & 0 & \cdots & 0 & \frac{1}{n_k}
\end{pmatrix}
\]

\( \Sigma = BAB' = \text{diag}(n_1^{-1} \lambda_{11}, n_2^{-1} \lambda_{22}, \ldots, n_k^{-1} \lambda_{kk}). \) Then \( \sigma_{ii} = n_i^{-1} \lambda_{ii} = m_i^{-1} \) and \( \sigma_{ij} = 0 \) for all \( i \neq j \). Let \( m^{-1}(i) = n_i^{-1} \lambda_{ii}(i), 1 \leq i \leq k \) and \( m[1] \leq \cdots \leq m[k] \). Then \( m_\ell^{-1} = (m_\ell^{-1}) + m_{(\ell)}^{-1} \), \( 1 \leq r, \ell \leq k, \ell \neq \ell; \) and for any \( \ell, 1 \leq \ell \leq k, \)
\[ \rho_{i,j}(\varepsilon) = \alpha_i \alpha_j \varepsilon_{m-1} = \frac{1}{(1 + \frac{m(i)}{m(j)})^{\frac{1}{2}}} (1 + \frac{m(i)}{m(j)})^{\frac{\varepsilon}{2}}, \]

for \( i \neq j, i, j \neq k \). Let

\[ (3.10) \quad \kappa_{i,j} = \frac{1}{(1 + \frac{m(i)}{m(j)})^{\frac{\varepsilon}{2}}} (1 + \frac{m(i)}{m(j)})^{\frac{\varepsilon}{2}}, \quad i \neq j, i, j \neq k. \]

Then, it is easy to check that the condition (3.8) is satisfied.

**Expected subset size for a special case**

Let the joint density \( p \) as in (3.2) have the form

\[ (3.11) \quad p(\chi) = h(\chi \Sigma^{-1} \chi) \]

where \( \Sigma = (\sigma_{i,j}) \) is positive definite with \( \sigma_{11} = \sigma_{22} = \ldots = \sigma_{kk} = \sigma^2 \) and \( \sigma_{ij} = \sigma^2 \rho \) when \( i \neq j \), \( \sigma \) and \( \rho \) are known. Let \( S \) be the size of the selected subset excluding the best population. Then the expected subset size is given by

\[ E(S|R) = \sum_{i=1}^{k-1} \text{P(Selecting } \pi_i \text{ } | R) \]

\[ = \sum_{i=1}^{k-1} \text{P}(Y(i) \geq \max_{1 \leq j \leq k} Y(j) - \text{Cov}(1-\rho)) \]

\[ = \sum_{i=1}^{k-1} \int_{B_{\varepsilon} + \theta_{\varepsilon}} a_k |(\rho_{i,j}(\varepsilon))|^{-\frac{3}{2}} f(z_{i,j}(\rho_{i,j}(\varepsilon))z_{i,j})dz_{i,j}, \]

where \( \theta_{\varepsilon} = (\theta_{[\varepsilon]} - \theta_{[r]})(2\sigma^2(1-\rho))^{-\frac{3}{2}}, r=1,2,\ldots,k, r \neq \varepsilon, \]

\[ B_{\varepsilon} + \theta_{\varepsilon} = \{ z_{r,k} \leq c(\theta_{[\varepsilon]} - \theta_{[r]})(2\sigma^2(1-\rho))^{-\frac{3}{2}}, \]

\[ r = 1,2,\ldots,k, r \neq \varepsilon, \quad 1 \leq \varepsilon \leq k, \]

and \( \rho_{i,j}(\varepsilon) \) defined as in (3.6) is
\[ \rho_{ij}(z) = \begin{cases} 1 & \text{if } i = j, i, j \neq \ell, \\ \frac{1}{2} & \text{if } i \neq j, i, j \neq \ell. \end{cases} \]

We assume that \( f \) is strictly decreasing. Then \( f \) is Schur-concave [8].

Since \( y \in B_\ell \) and \( x < y \) implies \( x \in B_\ell \), hence

\[ \int_{B_\ell + \theta_\ell} a_k |(\rho_{ij}(z))|^{-\frac{1}{2}} f(z^t(\rho_{ij}(z))^{-1} z) dz \]

is a Schur-concave function of \( \theta_\ell \) [10]. From the fact that \((a_1, a_2, \ldots, a_n) > (\frac{1}{n}, \ldots , 1)\) for all vectors \( a \), where \( a = (a_1, \ldots, a_n) \), \( b = (b_1, \ldots, b_n) \) and \( a_1 \geq a_2 \geq \ldots \geq a_n \), \( b_1 \geq b_2 \geq \ldots \geq b_n \), \( a > b \) means

\[ \sum_{i=1}^{\ell} a_i > \sum_{i=1}^{\ell} b_i, \; \ell = 1, \ldots, n-1, \; \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i. \]

For any \( \ell, 2 \leq \ell \leq k-1, \)

\[ (\theta[\ell] - \theta[1], \ldots, \theta[\ell] - \theta[1-\ell], \ldots, \theta[k] - \theta[k]) > (0, \ldots, 0), \] for some \( \theta \).

But \( \theta[\ell] - \theta[j] \leq 0 \) for \( j > \ell \) and \( \theta[\ell] - \theta[j] \geq 0 \) for \( j = \ell \), hence it follows that the supremum of

\[ \int_{B_\ell + \theta_\ell} a_k |(\rho_{ij}(z))|^{-\frac{1}{2}} f(z^t(\rho_{ij}(z))^{-1} z) dz \]

over \( \Omega \) occurs when \( \theta[1] = \ldots = \theta[k] \). For \( \ell = 1, \)

\[ (\theta[1] - \theta[2], \ldots, \theta[1] - \theta[k]) \leq (0, \ldots, 0) \]

and \( B_1 + \theta_1 \subset B_1 \). Hence

\[ \sup_{\theta_\ell \in \Omega} E_{\theta}(S|R) = \sum_{\ell=1}^{k-1} \int_{B_\ell} a_k |(\kappa_{ij}(z))|^{-\frac{1}{2}} f(z^t(\kappa_{ij}(z))^{-1} z) dz \]

\[ = (k-1)P^* \text{ provided that} \]

\[ \inf_{\theta_\ell \in \Omega} P_{\theta}(CS|R) = P^*. \]
Remark: Let \( p \) be the multivariate normal density as in (3.11), then \( f \) has the required property.

3.1. Applications to Normal Populations

Let \( \pi_1, \pi_2, \ldots, \pi_k \) be \( k \) independent normal populations with means \( \mu_1, \mu_2, \ldots, \mu_k \) and variances \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 \), respectively. Let \( \sigma_1^2 = \ldots = \sigma_k^2 = \sigma^2 \), where \( \sigma^2 \) may or may not be known.

Case (a): \( \sigma^2 \) known. We assume without any loss of generality that \( \sigma^2 = 1 \), and for this problem (3.2) assumes the following form:

\[
p(x) = (2\pi)^{-\frac{k}{2}} |D|^{-\frac{3}{2}} h((Y-\mu)'D^{-1}(Y-\mu)),
\]

where \( \mu' = (\mu_1, \ldots, \mu_k) \), \( h(x) = e^{-x} \), and \( D = \text{diag}(n_1^{-1}, \ldots, n_k^{-1}) \).

Gupta and Huang [6] proposed the following rule \( R_1 \) based on the sample means \( Y_i \) from \( \pi_i \), \( i = 1, 2, \ldots, k \).

\( R_1: \) Retain \( \pi_i \) in the selected subset if and only if

\[
Y_i \geq \max_{1 \leq j \leq k} \left( \frac{Y_j - c_1 \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} {\left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right),
\]

where \( c_1 = c_1(k, p, n_1, \ldots, n_k) > 0 \) is chosen so as to satisfy the \( P^* \)-condition.

For the condition (3.10), \( \lambda_{i1} = 1 \) implies \( m_i^{-1} = n_i^{-1} \), \( 1 \leq i \leq k \). Therefore any \( \ell, 1 \leq \ell \leq k \),

\[
\kappa_{ij}^2 = \left( 1 + \frac{n_i[k]}{n_i[j]} \right) \left( 1 + \frac{n_i[j]}{n_i[k]} \right)^{-\frac{3}{2}}, \quad i \neq j, \quad i, j \neq \ell,
\]

and

\[
\kappa_{ii}^2 = 1, \quad 1 \leq i \leq k, \quad i \neq \ell.
\]

Let \( \beta_{i} = \left( 1 + \frac{n_i[k]}{n_i[i]} \right)^{-\frac{3}{2}}, \quad i = 1, 2, \ldots, k-1 \). Thus \( \kappa_{ij}^k = \beta_i \beta_j \), \( i \neq j, \quad i, j = 1, \ldots, k-1 \) and \( \kappa_{ii}^k = 1, \quad 1 \leq i \leq k-1 \).
By (3.7), we have

\[(3.12) \quad \inf P(\text{CS}|R) \leq \frac{c_1}{c_1} = \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_1} \left(2\pi\right)^{-\frac{k}{2}} |\mathbf{Z}(\mathbf{k}, \mathbf{i}, \mathbf{j})|^{-\frac{k}{2}} f(\mathbf{Z}(\mathbf{k}, \mathbf{i}, \mathbf{j})^{-1}\mathbf{Z}_k) d\mathbf{z}_1 \cdots d\mathbf{z}_{k-1, k},\]

where \(Z_k, \ldots, Z_{k-1, k}\) are standard normal random variables with correlation \(K^{k}_{rs} = \beta^{k}_{r} \beta^{k}_{s}\). It is known that \(Z_k, \ldots, Z_{k-1, k}\) can be generated from \(k\) independent standard variates \(Y_1, \ldots, Y_{k-1, k}\) by the transformation

\[Z_{jk} = (1-\beta^{2}_{j})^{-\frac{1}{2}} Y_j + \beta^{2}_{j} Y,\]

and then (3.9) is as follows:

\[(3.13) \quad \inf P(\text{CS}|R_1) = \int \prod_{j=1}^{k-1} \phi\left(\frac{c_1-\beta^{2}_{j} u}{(1-\beta^{2}_{j})^{\frac{1}{2}}}\right) d\phi(u).\]

**Case (b):** \(\sigma^2\) unknown. Let \(s^2\) denote the usual pooled estimate of \(\sigma^2\) on \(\nu\) degrees of freedom. Gupta and Huang [6] proposed the rule \(R_2\) for selecting a subset containing the population associated with the largest \(\nu_i\)'s.

**\(R_2\):** Retain \(\pi_i\) in the selected subset if and only if

\[Y_i \geq \max_{1 \leq j < k} \left(\frac{Y_j - c_2 s^2\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}{n_i} \right),\]

where \(c_2 = c_2(k, P^*, n_1, \ldots, n_k) > 0\) is to be determined so that the \(P^*\)-condition is satisfied.

Using the same argument as in case (a), we can obtain

\[(3.14) \quad \inf P(\text{CS}|R_2) = \int \int \prod_{j=1}^{k-1} \phi\left(\frac{c_2 u - \beta^{2}_{j} x}{\sqrt{1-\beta^{2}_{j}}}\right) d\phi(x) dQ_\nu(u),\]

where \(Q_\nu(u)\) denotes the cdf of a \(\chi_\nu/\sqrt{\nu}\) variate.
The Evaluation of the Constant $c_1$ Associated with $R_1$

Let $U_1, \ldots, U_{k-1}$ be $k-1$ standard normal random variables, and the correlation coefficient of $U_i$ and $U_j$ be $\rho$, $i, j = 1, \ldots, k-1$, where

$$\rho = \left[ 1 + \frac{n[k]}{n[1]} \right] \left[ 1 + \frac{n[k]}{n[2]} \right]^{-\frac{1}{2}}.$$

Using the same notation as before, we have $\kappa_{ij}^k \geq \rho$, $i, j = 1, \ldots, k-1$, hence

$$P\left\{ Z_{ik} \leq c, i = 1, \ldots, k-1, \{\kappa_{ij}^k\} \right\} \geq P\{U_i \leq c, i = 1, \ldots, k-1, \{\rho\} \right\}$$

$$= \int_{-\infty}^{\infty} \phi^{k-1}(c-\rho \frac{\phi \sqrt{1-\rho}}{\sqrt{1-\rho}}) d\phi(u).$$

Equating the above integral to $P^*$, values of $c$ are available for the equi-correlated $U_i$'s from the tables in Gupta, Nagel and Panchapakesan [7]. These $c$-values will be greater than the exact $c_1$-values satisfying the equations by equating the left hand side of (3.13) to $P^*$. Some of the exact $c_1$-values can be obtained from Table 1 of Gupta and Huang [6].

Some Results on Comparisons

Assume that $\sigma^2 = 1$. The procedure of Gupta and D. Y. Huang [6] is more efficient than Gupta and W. T. Huang [5] for the case of $k = 2$,

$$n[1] = \alpha n[2], 0 < \alpha < 1.$$

For $\sigma^2$ unknown, Chen, Dudewicz and Lee [2], have proposed a class of procedures as follows:

$R_a$: Retain $Y_i$ in the selected subset if and only if

$$Y_i \geq \max_{1 \leq j \leq k} Y_j - q_a \sqrt{\frac{n_i}{\alpha}} + \sqrt{\frac{T}{n_i \alpha}}.$$
where $a$ is any fixed constant such that $0 < a < \infty$.

For any fixed $P^*$, $\frac{1}{k} < P^* < 1$, and $k = 2$, $n_1 + n_2$,

$$\inf_{\Omega} P(CS|R_a) = \int_{\Omega} \phi\left( \frac{\sqrt{n_1} + \frac{1}{a}}{\sqrt{n_2} + \frac{1}{a}} q_a x \right) dQ_y(x) = P^*,$$

and

$$\inf_{\Omega} P(CS|R_1) = \int_{\Omega} \phi(c_1 x) dQ_y(x) = P^*,$$

hence $c_1 \sqrt{\frac{n_1}{n_2} + \frac{1}{n_2}} = q_a \sqrt{\frac{n_1}{n_2} + \frac{1}{a}}$. Since

$$\sup_{\Omega} E(S|R_a) = \sup_{\Omega} \sum_{i=1}^{2} P(Y_i \geq \max_{1 \leq j \leq 2} Y_j - q_a s \sqrt{\frac{1}{n_1} + \frac{1}{a}}),$$

$$= \sup_{\Omega} \sum_{i=1}^{2} P(Y_i \geq \max_{1 \leq j \leq 2} Y_j - c_1 s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$$

$$= \sup_{\Omega} E(S|R_1).$$

4. Selection for Small Variances of Normal Populations

Let $\pi_1, \pi_2, \ldots, \pi_k$ denote $k$ independent normal populations with unknown variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively, ($\sigma_i > 0$, $i = 1, 2, \ldots, k$), and with all means known or unknown. The ordered variances are denoted by $\sigma^2_{[1]} \leq \cdots \leq \sigma^2_{[k]}$. It is assumed that there is no a priori information available about the correct pairing of the given populations and the ordered parameters $\sigma^2_{[i]}$. The population with variance equal to $\sigma^2_{[1]}$ is called the best population. The goal is to select a non empty subset of the $k$ populations containing the best population. Any such selection will be called a correct selection (CS).
Let \( s_1^2, s_2^2, \ldots, s_k^2 \) denote the sample variances. Let \( s_{(i)}^2 \) denote the (unknown) sample variance that is associated with the \( i \)th smallest population variance, \( \sigma_{[i]}^2 \); let \( \nu_{(i)} \) denote the number of degrees of freedom associated with \( s_{(i)}^2 \). Gupta and Sobel [8] have proposed a procedure for this goal. Gupta and Huang [6] obtained a lower bound on the infimum of the correct selection. We modify Gupta-Sobel procedure to obtain exact results to satisfy \( P^* \)-condition asymptotically and apply the method of Section 2 to obtain an optimal procedure.

\[ R_3: \text{ Retain } \pi_i \text{ in the selected subset if and only if } \]

\[
s_i^2 \leq \min_{1 \leq j < k} \left[ \left( \frac{1}{\nu_i} + \frac{1}{\nu_j} \right)^2 \sigma_{[i]}^2 \right],
\]

where \( c_3, (0 < c_3 < 1) \), is the largest value satisfying the basic \( P^* \)-condition.

We shall show how large sample theory can be used to find very good approximations to the required probabilities even for relatively small \( n \). Our principal tools will be the use of the transformation \( y = \log_3 s^2 \) (see [1]), and the approach of certain multivariate distributions to multivariate normal distributions.

Let \( X_i = \log_3 s_{\pi_i}^2, i = 1, 2, \ldots, k \). It is known (see [8]) that the expectation and variances are

\[
EX_i = -\left( \frac{1}{\nu_i} + \frac{1}{3\nu_i^2} \right) + O(\nu_i^{-3}) - \frac{1}{\nu_i},
\]

\[
\text{Var}(X_i) = \frac{d^2}{dx^2} \left[ \log_3 r(x) \right] \bigg|_{x=\frac{\nu_i}{2}} = \frac{2}{\nu_i} + \frac{2}{\nu_i^2} + \frac{4}{3\nu_i^3} + O(\nu_i^{-5}) - \frac{2}{\nu_i}, \quad i = 1, 2, \ldots, k,
\]
and
\[ E(X_i - X_j) = \frac{1}{\nu_j} - \frac{1}{\nu_i}, \]

\[ \text{Var}(X_i - X_j) = \frac{2}{\nu_i} + \frac{2}{\nu_j}, \text{ for } j = 1, 2, \ldots, k; j \neq i. \]

Thus
\[ \frac{2}{\nu_j} \log \frac{s_i^2}{\sigma_i^2} \]

is asymptotically distributed as a standard normal variable \( Z_i \) as \( \nu_i \to \infty \).

Since
\[ \frac{s_i^2}{\sigma_i^2} \leq \left( \frac{1}{c_3} \right) \left( \frac{1}{\nu_j} + \frac{1}{\nu_j} \right)^{\beta} \frac{s_i^2}{\sigma_j^2}, \text{ for } j = 1, 2, \ldots, k, j \neq i \]

is equivalent to
\[ Z_{ij} \leq \frac{1}{\sqrt{2}} \log \frac{1}{c_3} + \frac{1}{\nu_i} - \frac{1}{\nu_j} \sqrt{\frac{2}{\nu_i} + \frac{2}{\nu_j}}, \text{ for } j = 1, 2, \ldots, k, j \neq i, \]

where
\[ Z_{ij} = (X_i - X_j + \frac{1}{\nu_i} - \frac{1}{\nu_j})(\frac{2}{\nu_i} + \frac{2}{\nu_j})^{-\frac{\beta}{2}}, \text{ for } 1 \leq i, j \leq k, i \neq j, \text{ and } \]

\[ \kappa_{ij}^k = \frac{1}{\sqrt{1 + \frac{\nu[k]}{\nu[i]}(1 + \frac{\nu[k]}{\nu[j]})}} = r_{ik} r_{jk}, i, j = 1, \ldots, k-1, i \neq j. \]

Hence we can apply the results in Section 3 to prove
\[ \inf P(CS|R_3) \approx \prod_{j=1}^{k-1} \phi \left( \frac{c_{kj} - r_{jk} u}{(1 - r_{kj}^2)^{\frac{\beta}{2}}} \right) d\phi(u), \]

where
\[ c_{kj} = \frac{1}{\sqrt{2}} \log \frac{1}{c_3} + \frac{1}{\nu_i} - \frac{1}{\nu_j} \left( \frac{\nu[k]}{\nu[i]} + \frac{\nu[j]}{\nu[j]} \right)^{\frac{\beta}{2}}, \text{ for } j = 1, \ldots, k-1. \]
It should be pointed out that for equal sample size case, Gupta and
Sobel [8] compared the exact value and the asymptotic values of the constant
c3 to see how close they are.

For any i, 1 ≤ i ≤ k, let \( \tau_{ij} = \frac{s^2}{\sigma_i^2} \) and \( T_{ij} = \frac{s^2}{\sigma_i^2} \) for 1 ≤ j ≤ k, j ≠ i.

We can find the joint density of \( T_{ij}, j = 1,2,...,k, j \neq i \). We can construct
an optimal procedure based on \( T_{ij} \)'s using the method of Section 2.

5. Selection Procedures in Two-Way Layouts

Let \( \pi_1, \pi_2, ..., \pi_k \) be k populations. For a two-factor complete block
design with one observation per cell, we express the observable random
variables \( X_{i\alpha} \) (i = 1,2,...,k; \( \alpha = 1,...,n \)) as

\[
X_{i\alpha} = \mu + \beta_{\alpha} + \tau_i + \xi_{i\alpha}, \quad \sum_{i=1}^{k} \tau_i = 0,
\]

where \( \mu \) is the mean-effect, \( \beta_1, ..., \beta_{\alpha} \) are the block effects (nuisance
parameters for the fixed effects model or random variables for the mixed
effects model), \( \tau_1, ..., \tau_k \) are the treatment effects, and the \( \xi_{i\alpha} \) are the
error components. Let \( X_{i1}, ..., X_{in} \) denote the n independent observations
from the ith population \( \pi_i \). Let the joint density of \( X_{11}, ..., X_{1n}; X_{21}, ..., X_{2k}; ..., X_{k1}, ..., X_{kn} \) be of the following form:

\[
c_k |\Lambda|^{-\frac{3}{2}} g(x-\theta) \Lambda^{-1}(x-\theta)
\]

where \( x' = (x_{11}, ..., x_{1n}; ..., x_{k1}, ..., x_{kn}) \), and \( \theta' = (\theta_{11}, ..., \theta_{1n}; ..., \theta_{k1}, ..., \theta_{kn}) \),
\( \theta_{i\alpha} = \mu + \beta_{\alpha} + \tau_i, \ i = 1,2,...,k; \ \alpha = 1,...,n, \) and \( \Lambda \) is a known positive
definite matrix, \( c_k \) is determined such that (5.2) is a density.

Our purpose is to study some selection procedures to select a subset
of a random size containing the "best" treatment. The quality of the
treatment is judge by the largeness of the $\tau_1$'s.

Let $\tau_1 \leq \ldots \leq \tau_k$ be the actual ranked $\tau$'s (which are unknown), and let

$$Z_i = X_i - \bar{X} \text{ where } X_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}, \ i = 1, \ldots, k; \text{ and } \bar{X} = \frac{1}{k} \sum_{i=1}^{k} X_i.$$

We denote the ordered values of the $Z_i$'s by $Z_{[1]} \leq \ldots \leq Z_{[k]}$ and let $Z_{(i)}$ be the random variable associated with $\tau_{[i]}$, $i = 1, \ldots, k$.

By a similar argument as in Section 3, we know that $Z_i$ is the generalized least square estimator of $\tau_i$.

Let $Z = EY$, where $E' = (E_1, \ldots, E_k)'_{kn 	imes k}$ with rank $k$, $Y' = (X_{11}, \ldots, X_{1n}; \ldots; X_{k1}, \ldots, X_{kn})$, and $E_i' = (-\frac{1}{kn}, \ldots, -\frac{1}{kn}; \ldots; \frac{1}{n} (1 - \frac{1}{k}), \ldots, \frac{1}{n} (1 - \frac{1}{k}); \ldots; -\frac{1}{kn}, \ldots, -\frac{1}{kn})$, for $1 \leq i \leq k$.

Then the joint density of $Z_1, \ldots, Z_k$ is of the form:

$$b_k |\Sigma|^{-\frac{k}{2}} h((z-\bar{z})'\Sigma^{-1}(z-\bar{z}))$$

where $z' = (z_1, \ldots, z_k)$, $z' = (\tau_1, \ldots, \tau_k)$ and $\sum_{k \times k} = E \wedge E' = (\sigma_{ij})$.

The methods to construct selection procedures are the same as in Section 3.

References


Some Multiple Decision Problems in Analysis of Variance

Shanti S. Gupta and D. Y. Huang

Purdue University
Department of Statistics
W. Lafayette, IN 47907

Office of Naval Research
Washington, DC

Approved for public release, distribution unlimited.

Locally best, selection procedures, correct selection, generalized LS estimates, Schur-concave functions, unequal sample sizes.
some other inferences about the parameters of interest.

The extensive literature on selection and ranking procedures depends heavily on the use of independence between populations (block, treatments, etc.) in the analysis of variance. In practical applications, it is desirable to drop this assumption of independence and consider cases more general than the normal.

In the present paper, we derive a method to construct locally best (in some sense) selection procedures to select a non empty subset of the k populations containing the best population as ranked in terms of $\theta_i$'s which control the size of the selected subset and which maximizes the probability of selecting the best. We also consider the usual selection procedures in one-way ANOVA based on the generalized least squares estimates and apply the method to two-way layout case. Some examples are discussed and some results on comparisons with other procedures are also obtained.