AD-A101 922 PURDUE UNIV LAFAYETTE IN DEPT OF STATISTICS F/G 12/1
SELECTION PROCEDURES FOR A PROBLEM IN ANALYSIS OF VARIANCE.(U)
JUN 81 S S GUPTA, D HUANG N00014-75-C-0455
UNCLASSIFIED MMS-81-22 NL

1 OF 1
AD A
101 922

END
DATE
FILMED
8-81
DTIC

LEVEL II

(12)

AD A101922

# PURDUE UNIVERSITY



purdue university

## DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES





81 7 21 025

12

Selection Procedures for A
Problem in Analysis of Variance*

by

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

9 Mimeograph Series #81-22

14) MMS-81-22

12 49



Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #81-22

11) June 1981

15) *This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.



21 73 ... 01

# Selection Procedures For A Problem in Analysis of Variance*

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

## 1. Introduction

For a completely randomized block design with one observation per cell, we express the observable random variables $X_{i\alpha}$ $(i = 1,\ldots,k;$ $\alpha = 1,\ldots,n)$ as

$$(1.1)\qquad X_{i\alpha} = \mu + \beta_\alpha + \tau_i + \epsilon_{i\alpha},\quad \sum_{i=1}^{k} \tau_i = 0,$$

where $\mu$ is the mean-effect, $\beta_1,\ldots,\beta_n$ are the block effects (nuisance parameters for the fixed effects model), $\tau_1,\ldots,\tau_k$ are the treatment effects, and $\epsilon_{i\alpha}$ are the error components. We assume that the errors within each block are jointly normally distributed.

We assume that the quality of a treatment is judged by the largeness of the $\tau_i$'s. A 'population' $\pi_i$ is called the best if $\tau_i$ is the largest. In general, it may be complicated to derive suitable tests for appropriate hypotheses, in which the experimenter may really be interested. We apply the subset selection approach (using certain basic hypotheses) and thus obtain more appropriate information regarding the treatments. A subset selection procedure is designed to select a subset so as to include the best population. Selection of any subset that contains the best is called a correct selection (CS). Roughly speaking, any two populations that are in the same selected subset, will be considered as "equivalently good". If all populations are selected, we claim that all treatments are homogeneous. In general, for achieving the objective of the experimenter, one should establish a suitable set of basic hypotheses. Depending on the objective one should proceed to consider different

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ways of formulating the basic hypotheses. In this paper, we discuss a method based on subset selection rules for the purpose of making a claim of the type: $\tau_i = \tau^* > \tau_j + \Delta$ for all $i \in I$ and $j \in J$, where I and J form a partition of $\{1,2,\ldots,k\}$. The process of making such a claim will be called hypothesis identification. This is achieved by setting up certain basic hypotheses regarding the $\tau_i$'s and using a subset selection procedure to test these basic hypotheses. It should be pointed out that in identifying an appropriate hypothesis, we assume that the constant $\Delta$ in the claim is specified by the experimenter, say, based on past experience. Associated with the tests of the basic hypotheses using a selection rule, there are error probabilities and the infimum of the probability of a correct selection for the rule employed. These are related to the power function of these tests. The sum of the average (over the basic hypotheses tested) of the error probabilities and one minus the infimum of the probability of a correct selection is called the identification risk. The main theorem of the paper discusses the derivation of an optimal selection rule in the sense of minimizing the identification risk. For a more general theory of multiple decisions from ranking and selection approach, one can refer to a recent monograph by Gupta and Huang (1981). A general survey of the entire field is provided in Gupta and Panchapakesan (1979).

Let $\underline{Y}$ be a random observable vector with probability distribution depending upon a parameter $\underline{\tau}' = (\tau_1,\ldots,\tau_k) \in \Omega$. Consider a family of hypotheses testing problems as follows:

(1.2) $$H_0:\ \underline{\tau} \in \Omega_0 \quad \text{vs} \quad H_i:\ \underline{\tau} \in \Omega_i,\ 1 \le i \le k,$$

where $\Omega_0 = \{\underline{\tau} | \tau_1 = \ldots = \tau_k\}$ and $\Omega_i = \{\underline{\tau} | \tau_i > \max_{j \neq i} \tau_j\}$, $i = 1,2,\ldots,k$. A test of the hypotheses (1.2) will be defined to be a vector $(\delta_1(\underline{y}),\ldots,\delta_k(\underline{y}))$, where the elements of the vector are ordinary test functions; when $\underline{y}$ is observed we reject $H_0$ in favor of $H_i$ with probability $\delta_i(\underline{y})$, $1 \leq i \leq k$. The power function of a test $(\delta_1,\ldots,\delta_k)$ is defined to be the vector $(\beta_1(\underline{\tau}),\ldots,\beta_k(\underline{\tau}))$, where $\beta_i(\underline{\tau}) = E_{\underline{\tau}}\delta_i(\underline{Y})$, $1 \leq i \leq k$. For $\underline{\tau} \in \Omega_i$, $\beta_i(\underline{\tau})$ is the probability of a correct selection P(CS) and $\delta_i(\underline{y})$ is the individual selection probability of selecting the best population $\pi_i$. Let $S_\gamma$ be the set of all the tests $(\delta_1,\ldots,\delta_k)$ such that

$$(1.3) \qquad E_{\underline{\tau}}\delta_i(\underline{Y}) \leq \gamma, \quad \underline{\tau} \in \Omega_0, \quad 1 \leq i \leq k,$$

where $\gamma$ is the upper bound on the error probabilities associated with the treatment effects.

For each i, $(1 \leq i \leq k)$, we would like to have $\beta_i(\underline{\tau})$ large when $\underline{\tau} \in \Omega_i$ subject to (1.3). For $\underline{\tau} \in \Omega_i$, if we make $\beta_i(\underline{\tau})$ large, then $\beta_j(\underline{\tau})$ should be small for $j \neq i$.

It should be pointed out that in the formulation and proof of the optimal selection procedure, results from Neyman-Pearson theory are used.

## 2. Formulation of an Optimal Selection Procedure

Assume that

$$\underline{X}_\alpha' = (X_{1\alpha},\ldots,X_{k\alpha}),$$

$\alpha = 1,\ldots,n$, are independently and identically distributed random vectors with the following distribution:

$$(2.1) \qquad (2\pi\sigma^2)^{-\frac{1}{2}kn}|\Lambda|^{-\frac{1}{2}}\exp[-\frac{1}{2\sigma^2}(\underline{x}-\underline{\theta})'\Lambda^{-1}(\underline{x}-\underline{\theta})],$$

where $\underline{x}' = (x_{11},\ldots,x_{k1};\ldots;x_{1n},\ldots,x_{kn})$ and $\underline{\theta}' = (\theta_{11},\ldots,\theta_{k1};\ldots;\theta_{1n},\ldots,\theta_{kn})$, $\theta_{i\alpha} = \mu + \beta_\alpha + \tau_i$, $i = 1,\ldots,k$; $\alpha = 1,2,\ldots,n$ and $\Lambda$ is a known positive definite kn x kn correlation matrix defined as follows:

$\Lambda = (\lambda_{ij})_{kn \times kn}$

$$= \begin{bmatrix} \Lambda_1 & 0 & \dots & 0 \\ 0 & \Lambda_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \Lambda_1 \end{bmatrix}, \quad \text{where}$$

$$\Lambda_1 = \begin{bmatrix} 1 & & \lambda \\ & \ddots & \\ \lambda & & 1 \end{bmatrix}_{k \times k}.$$

We rewrite the original model as the general linear model as follows:

$$\underline{X} = \underline{\theta} + \underline{\epsilon}, \qquad \underline{\epsilon} \sim N(\underline{0}, \sigma^2 \Lambda).$$

Since we are interested in the difference between all pairs of $\tau_i$'s, we transform the linear model to the following: For any i, let

$$\underline{Y}_i = C\, \underline{\tau}_i + \underline{\eta}, \qquad \underline{\eta} \sim N(\underline{0}, \sigma^2 \Sigma_i),$$

where $\underline{\tau}_i' = (\tau_{i1}, \dots, \tau_{ik})$, $\tau_{ij} = \tau_i - \tau_j$, $j \neq i$,

$\underline{Y}_i' = (Y_{i11}, \dots, Y_{ik1}; \dots; Y_{i1n}, \dots, Y_{ikn})_{1 \times (k-1)n}$

$Y_{ij\ell} = X_{i\ell} - X_{j\ell}$, $i \neq j$; $i,j = 1, \dots, k$; $\ell = 1, \dots, n$,

$\underline{Y}_i = A_i\, \underline{X}$, $\qquad \underline{\eta} = A_i\, \underline{\epsilon}$

$$A_i = \begin{bmatrix} A_{i1} & & & \\ & A_{i1} & & 0 \\ & 0 & \ddots & \\ & & & A_{i1} \end{bmatrix}_{(k-1)n \times kn}$$

$$\Sigma_i = A_i \Lambda A_i' = \begin{bmatrix} A_{i1}\Lambda_1 A_{i1}' & & 0 \\ & \ddots & \\ 0 & & A_{i1}\Lambda_1 A_{i1}' \end{bmatrix}_{(k-1)n\times(k-1)n},$$

$$A_{i1} = \begin{array}{c} \begin{array}{ccccc} & & \quad\quad\quad\quad\quad\quad\quad\quad i & & \end{array} \\ \left[\begin{array}{cccccccc} -1 & 0\cdots\cdots 0 & & & 1 & 0\cdots\cdots\cdots 0 & & \\ 0 & -1 & 0\cdots 0 & & 1 & 0\cdots\cdots\cdots 0 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots\cdots 0 & & & -1 & 1 & 0\cdots\cdots\cdots 0 & & \\ 0\cdots\cdots\cdots 0 & & & & 1 & -1 & 0\cdots\cdots 0 & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots\cdots\cdots & & & & 1 & 0\cdots\cdots 0 & & -1 \end{array}\right] \begin{array}{l} \\ \\ \\ i-1 \\ i+1 \\ \\ (k-1)\times k \end{array} \end{array}$$

$$C' = [I,\ldots,I]_{(k-1)\times(k-1)n}$$

where each of the identity matrix in $C'$ is $(k-1)\times(k-1)$.
The maximum likelihood estimator of $\underline{\tau}_i$ is as follows:

$$\hat{\underline{\tau}}_i = (C'\Sigma_i^{-1}C)^{-1}C'\Sigma_i^{-1}\underline{Y}_i.$$

Since,

$$A_{i1}\ \Lambda_1\ A_{i1}' = (1-\lambda) \begin{bmatrix} 2 & & 1 \\ & \ddots & \\ 1 & & 2 \end{bmatrix}_{(k-1)\times(k-1)}$$

$$(A_{i1}\ \Lambda_1\ A_{i1}')^{-1} = (1-\lambda)^{-1}\frac{1}{k} \begin{bmatrix} k-1 & & -1 \\ & \ddots & \\ -1 & & k-1 \end{bmatrix} = V_i$$

$$C'\Sigma_i^{-1}C = n\ (A_{i1}\ \Lambda_1\ A_{i1}')^{-1} = \frac{n}{k(1-\lambda)} \begin{bmatrix} k-1 & & -1 \\ & \ddots & \\ -1 & & k-1 \end{bmatrix}$$

$$(C'\Sigma_i^{-1}C)^{-1} = \frac{1-\lambda}{n} \begin{bmatrix} 2 & & 1 \\ & \ddots & \\ 1 & & 2 \end{bmatrix}_{(k-1)\times(k-1)},$$

$$C'\Sigma_i^{-1} = [I\ldots I]\begin{bmatrix} V_i & & 0 \\ & \ddots & \\ 0 & & V_i \end{bmatrix}$$

$$= [V_i,\ldots,V_i]$$

$$(C'\Sigma_i^{-1}C)^{-1}C'\Sigma_i^{-1} = \frac{1-\lambda}{n} \begin{bmatrix} 2 & & 1 \\ & \ddots & \\ 1 & & 2 \end{bmatrix} [V_i,\ldots,V_i]$$

$$= \frac{1}{n}[I,\ldots,I].$$

Hence,

$$\hat{\underline{\tau}}_i = (C'\Sigma_i^{-1}C)^{-1}C'\Sigma_i^{-1}\,\underline{Y}_i$$

$$= \frac{1}{n}\begin{bmatrix} \sum_{\ell=1}^{n} Y_{i1\ell} \\ \vdots \\ \sum_{\ell=1}^{n} Y_{ik\ell} \end{bmatrix} = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{ik} \end{bmatrix} = \begin{bmatrix} \bar{X}_i - \bar{X}_1 \\ \vdots \\ \bar{X}_i - \bar{X}_k \end{bmatrix},$$

where $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$, $1 \le i \le k$.

The joint density of $Y_{i11},\ldots,Y_{ik1};\ldots;Y_{i1n},\ldots,Y_{ikn}$ is the following:

$$p_{\underline{\tau}_i}(\underline{y}_i) = (2\pi\sigma^2)^{-\frac{1}{2}k}|\Sigma_i|^{-\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\underline{y}_i - C\underline{\tau}_i)'\Sigma_i^{-1}(\underline{y}_i - C\underline{\tau}_i)]$$

where

$$\Sigma_i = A_i \,\Lambda\, A_i' = (1-\lambda)\begin{bmatrix} J & & 0 \\ & \ddots & \\ 0 & & J \end{bmatrix}_{(k-1)n\times(k-1)n}$$

$$J = \begin{bmatrix} 2 & & 0 \\ & \ddots & \\ 0 & & 2 \end{bmatrix}_{(k-1)\times(k-1)}.$$

$$\Sigma_i^{-1} = \begin{bmatrix} V_i & & 0 \\ & \ddots & \\ 0 & & V_i \end{bmatrix}.$$

Now, we specify the $\Omega_i$'s as follows (Note that this is a different specification from that given earlier):

$$\Omega_i = \{\underline{\tau} | \tau_i \geq \max_{j \neq i} \tau_j + \Delta\sigma\}, \quad 1 \leq i \leq k,$$

and

$$\bar{\Omega} = \bigcup_{i=1}^{k} \Omega_i.$$

Assume that $\sigma$ is known. Let

$$\underline{\Delta}_i' = (\Delta\sigma, \ldots, \Delta\sigma)_{1 \times (k-1)}, \quad i = 1, \ldots, k, \ \Delta > 0.$$

Thus

$$\frac{p_{\underline{\Delta}_i}(\underline{y}_i)}{p_{\underline{0}}(\underline{y}_i)} = \exp \frac{1}{2\sigma^2} \{-(\underline{y}_i - C\underline{\Delta}_i)' \Sigma_i^{-1} (\underline{y}_i - C\underline{\Delta}_i) + \underline{y}_i' \Sigma_i^{-1} \underline{y}_i\}$$

$$= \exp\{\frac{1}{\sigma^2} \underline{\Delta}_i' C' \Sigma_i^{-1} \underline{y}_i - \frac{1}{2\sigma^2} \underline{\Delta}_i' C' \Sigma_i^{-1} C \, \underline{\Delta}_i\}$$

$$= \exp\{\frac{n\Delta}{(1-\lambda)k\sigma}(y_{i1} + \ldots + y_{ik}) - \frac{1}{2\sigma^2} \underline{\Delta}_i' C' \Sigma_i^{-1} C \underline{\Delta}_i\}.$$

Hence, we can rewrite

$$\frac{p_{\underline{\Delta}_i}(\underline{y}_i)}{p_{\underline{0}}(\underline{y}_i)} \geq d' \quad \text{as}$$

$$y_{i1} + \ldots + y_{ik} \geq d''\sigma.$$

Let a selection rule $\delta^0 = (\delta_1^0, \ldots, \delta_k^0)$ be defined by

$$\delta_i^0(\underline{y}_i) = \begin{cases} 1 & \text{if } p_{\underline{\Delta}_i}(\underline{y}_i) \geq d' p_{\underline{0}}(\underline{y}_i), \\ 0 & \text{otherwise} \end{cases},$$

such that

(2.2) $\quad E_{\underline{\tau}} \delta^0(\underline{Y}_i) = \gamma, \ \underline{\tau} \in \Omega_0.$ Then

$\delta^0$ maximizes

(2.3) $$\inf_{\bar{\Omega}} P(CS|\delta)$$

among all selection rules $\delta \in S(\gamma)$.

Note that $\delta_i^0(\underline{y}_i)$ is also based on the maximum likelihood estimators $\hat{\underline{\tau}}_i$ of $\underline{\tau}_i$. Since for any $\delta \in S(\gamma)$,

$$\underline{\tau} \in \bar{\Omega} = \bigcup_{i=1}^{k} \Omega_i \text{ implies } \underline{\tau} \in \Omega_i \text{ for some } i, \text{ thus}$$

$$P(CS|\delta) = \int \delta_i(\underline{y}_i) p_{\underline{\tau}}(\underline{y}_i) d\nu(\underline{y}_i)$$

$$\geq \min_{1 \leq i \leq k} \inf_{\underline{\tau} \in \Omega_i} \int \delta_i(\underline{y}_i) p_{\underline{\tau}}(\underline{y}_i) d\nu(\underline{y}_i).$$

We have

$$\inf_{\underline{\tau} \in \bar{\Omega}} P(CS|\delta) = \min_{1 \leq i \leq k} \inf_{\underline{\tau} \in \Omega_i} \int \delta_i(\underline{y}_i) p_{\underline{\tau}}(\underline{y}_i) d\nu(\underline{y}_i).$$

For any $\delta \in S(\gamma)$, it follows that

$$\int (\delta_i - \delta_i^0)(p_{\underline{\Delta}_i} - dp_{\underline{0}}) \leq 0$$

which implies

$$\int \delta_i^0 \, p_{\underline{\Delta}_i} \geq \int \delta_i p_{\underline{\Delta}_i}.$$

Since $\delta_i^0(\underline{y}_i)$ is nondecreasing in $\underline{y}_i$, hence

$$\inf_{\underline{\tau} \in \bar{\Omega}} P(CS|\delta^0) = \min_{1 \leq i \leq k} \int \delta_i^0(\underline{y}_i) p_{\underline{\Delta}_i}(\underline{y}_i) d\nu(\underline{y}_i)$$

$$\geq \min_{1 \leq i \leq k} \int \delta_i(\underline{y}_i) p_{\underline{\Delta}_i}(\underline{y}_i) d\nu(\underline{y}_i)$$

$$\geq \min_{1 \leq i \leq k} \inf_{\underline{\tau} \in \Omega_i} \int \delta_i(\underline{y}_i) p_{\underline{\tau}}(\underline{y}_i) d\nu(\underline{y}_i)$$

$$= \inf_{\underline{\theta} \in \bar{\Omega}} P(CS|\delta).$$

We rewrite $\delta^0$ as follows:

$$\delta_i^0(\underline{y}_i) = \begin{cases} 1 & \text{if } \quad y_{i1} + \ldots + y_{ik} \geq d''\sigma, \\ 0 & \text{otherwise} \end{cases}$$

Thus, the optimal subset selection rule is as follows:

$$\delta_i^0(\underline{x}) = \begin{cases} 1 & \text{if } \bar{x}_i \geq \frac{1}{k-1} \sum_{j \neq i} \bar{x}_j + d\sigma, \\ 0 & \text{otherwise} \end{cases},$$

where $d = \frac{d''}{k-1}$.

Now, we wish to determine d and n. We make the following transformation:

$$z_{ik} = [1 \ldots 1]_{1\times(k-1)} \begin{bmatrix} y_{i1} \\ \vdots \\ y_{ik} \end{bmatrix}, \quad \text{and}$$

$$\tau = \tau_{i1} + \ldots + \tau_{ik} = (k-1)\tau_i - \sum_{j \neq i} \tau_j.$$

Since the distribution of

$$\hat{\underline{\tau}}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{ik} \end{bmatrix} = (C'\Sigma_i^{-1}C)^{-1}C'\,\Sigma_i^{-1}\underline{Y}_i$$

is $(2\pi\sigma^2)^{-\frac{1}{2}k}|\Sigma_{1i}|^{-\frac{1}{2}}\exp[-\frac{1}{2\sigma^2}(\hat{\underline{\tau}}_i - \underline{\tau}_i)'\Sigma_{1i}^{-1}(\hat{\underline{\tau}}_i - \underline{\tau}_i)]$, where $\Sigma_{1i} = \frac{1-\lambda}{n} J$.

Then the distribution of $Z_{ik}$ is

$$[2\pi\sigma^2(1-\lambda)k(k-1)\,\frac{1}{n}]^{-\frac{1}{2}}\exp[-\frac{n}{2\sigma^2(1-\lambda)k(k-1)}(z_{ik}-\tau)^2].$$

Hence

$$E_{\underline{0}}\delta_i^0(\underline{Y}_i) = P(Z_{ik} \geq d''\sigma)$$

(2.4) $$= \Phi[-\frac{d''\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}}] = \gamma,$$

and

$$\inf_{\underline{\tau}\in\bar{\Omega}} P_{\underline{\tau}}(CS|\delta^0)$$

$$= \min_{1\le i\le k} \int \delta_i^0(\underline{y}_i)p_{\underline{\Delta}_i}(\underline{y}_i)d\nu(\underline{y}_i)$$

$$= \min_{1\le i\le k} P_{\underline{\Delta}_i}(Z_{ik} \ge d''\sigma)$$

$$= \min_{1\le i\le k} P_{\underline{\Delta}_i}\left(\frac{(Z_{ik}-(k-1)\Delta)\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}} \ge \frac{(d''-(k-1)\Delta)\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}}\right)$$

$$(2.5) \qquad = \Phi\left[-\frac{(d''-(k-1)\Delta)\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}}\right] = P^*.$$

For given r, $P^*$, k, $\lambda$, and $\Delta$, we can find d" and the smallest number of blocks, n, to satisfy equations (2.4) and (2.5). Note that this n is also the minimum sample size for the case of one observation per cell in the completely randomized block design.

We rewrite (2.4) and (2.5) as

$$\Phi\left[-\frac{d\sqrt{n(k-1)}}{\sqrt{(1-\lambda)k}}\right] = \gamma$$

and

$$\Phi\left[-\frac{(d-\Delta)\sqrt{n(k-1)}}{\sqrt{(1-\lambda)k}}\right] = P^*.$$

Let $z_{P*}$ and $z_\gamma$ represent the upper percentage points corresponding to $P^*$ and $\gamma$, respectively of the standard normal distribution. Then we have

$$d = -\frac{z_\gamma\Delta}{z_{P*} - z_\gamma},$$

and

$$n = \left\langle\frac{(1-\lambda)k(z_{P*} - z_\gamma)^2}{(k-1)\Delta^2}\right\rangle,$$

where <a> is the smallest integer greater than or equal to a.

Summarizing the previous results, we obtain the following theorem.

Theorem: Under model (1.1) with the stated assumption on $\underset{\sim}{\epsilon}_\alpha$, an optimal procedure for selecting a subset of the "best" or "worthwhile" treatments based on the observed data $\underset{\sim}{x}$ and satisfying the conditions (2.2) and (2.3) is: Select the population $\pi_i$ with probability $\delta_i^0(\underset{\sim}{x})$ given by

$$\delta_i^0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \bar{x}_i \geq \frac{1}{k-1} \sum_{j \neq i} \bar{x}_j + d\sigma, \\ 0 & \text{otherwise} \end{cases},$$

where the smallest values of d and n are given by

$$d = -\frac{z_\gamma \Lambda}{z_{p*} - z_\gamma},$$

and

$$n = \left\langle \frac{(1-\lambda)k(z_{p*} - z_\gamma)^2}{(k-1)\Delta^2} \right\rangle.$$

Furthermore, we have established the following connection between the selection procedure and the hypothesis identification problem as follows: If $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_j}$ $(j < k)$ are selected, we say that these populations are homogeneous and make the hypothesis identification

$$H_i': \tau_{i_1} = \ldots = \tau_{i_j} \geq \max_{\substack{1 \leq \ell \leq k \\ \ell \notin \{i_1, \ldots, i_j\}}} \tau_\ell + \Delta\sigma.$$

Note that the overall identification risk connected with this problem is $\leq \gamma + (1-P^*)$.

Remark: It should be pointed out that for some pairs $(\gamma, P^*)$, $\delta^0$ may not select any population. This is to be interpreted as not identifying any one of the appropriate hypotheses.

We consider some special cases to provide an idea as to the appropriate identification of one of the hypotheses. For $\gamma = 0.05, \lambda = 0.5$ and $P^* = 0.95, 0.90, 0.80$; then

(i) $k = 2$,

$$H_0: \tau_1 = \tau_2, \; H_1': \tau_1 \geq \tau_2 + \Delta\sigma, \; H_2': \tau_2 \geq \tau_1 + \Delta\sigma.$$

In this case, for specified $\Delta$-values, the smallest d and n needed for the optimal selection rule are given in the following table.

| $\Delta$ | 0.1 | 0.5 | 1.0 | 2.0 |
|---|---|---|---|---|
| d(0.95,0.90,0.80) | 0.05,0.06,0.07 | 0.25,0.32,0.33 | 0.50,0.64,0.66 | 1.00,1.29,1.33 |
| n(0.95,0.90,0.80) | 1089,858,620 | 44,35,25 | 11,9,7 | 3,3,2 |

(ii) $k = 3$,

$H_0: \tau_1 = \tau_2 = \tau_3$, $\quad H_1': \tau_1 \geq \max(\tau_2, \tau_3) + \Delta\sigma$,

$H_2': \tau_2 \geq \max(\tau_1, \tau_3) + \Delta\sigma$, $\quad H_3': \tau_3 \geq \max(\tau_1, \tau_2) + \Delta\sigma$,

$H_4': \tau_1 = \tau_2 \geq \tau_3 + \Delta\sigma$, $\quad H_5': \tau_1 = \tau_3 \geq \tau_2 + \Delta\sigma$,

$H_6': \tau_2 = \tau_3 \geq \tau_1 + \Delta\sigma$.

For optimal selection rule, the minimum value of d and n are computed (for specified values of $\Delta$) and given in the following table.

| $\Delta$ | 0.1 | 0.5 | 1.0 | 2.0 |
|---|---|---|---|---|
| d(0.95,0.90,0.80) | 0.05,0.06,0.07 | 0.25,0.32,0.33 | 0.50,0.64,0.66 | 1.00,1.29,1.33 |
| n(0.95,0.90,0.80) | 817,644,465 | 33,26,19 | 9,7,5 | 3,2,2 |

(iii) $k = 4$,

$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4$, $\quad H_1': \tau_1 \geq \max(\tau_2, \tau_3, \tau_4) + \Delta\sigma$,

$H_2'$: $\tau_2 \geq \max(\tau_1,\tau_3,\tau_4) + \Delta\sigma$, $H_3'$: $\tau_3 \geq \max(\tau_1,\tau_2,\tau_4) + \Delta\sigma$,

$H_4'$: $\tau_4 \geq \max(\tau_1,\tau_2,\tau_3) + \Delta\sigma$, $H_5'$: $\tau_1 = \tau_2 \geq \max(\tau_3,\tau_4) + \Delta\sigma$,

$H_6'$: $\tau_1 = \tau_3 \geq \max(\tau_2,\tau_4) + \Delta\sigma$, $H_7'$: $\tau_1 = \tau_4 \geq \max(\tau_2,\tau_3) + \Delta\sigma$,

$H_8'$: $\tau_2 = \tau_3 \geq \max(\tau_1,\tau_4) + \Delta\sigma$, $H_9'$: $\tau_2 = \tau_4 \geq \max(\tau_1,\tau_3) + \Delta\sigma$,

$H_{10}'$: $\tau_3 = \tau_4 \geq \max(\tau_1,\tau_2) + \Delta\sigma$, $H_{11}'$: $\tau_1 = \tau_2 = \tau_3 \geq \tau_4 + \Delta\sigma$,

$H_{12}'$: $\tau_1 = \tau_2 = \tau_4 \geq \tau_3 + \Delta\sigma$, $H_{13}'$: $\tau_1 = \tau_3 = \tau_4 \geq \tau_1 + \Delta\sigma$,

$H_{14}'$: $\tau_2 = \tau_3 = \tau_4 \geq \tau_1 + \Delta\sigma$.

For the optimal selection rule, the minimum value of d and n are computed (for specified values of $\Delta$) and given in the following table.

| $\Delta$ | 0.1 | 0.5 | 1.0 | 2.0 |
|---|---|---|---|---|
| d(0.95,0.90,0.80) | 0.05,0.06,0.07 | 0.25, 0.32, 0.33 | 0.50,0.64,0.66 | 1.00,1.29,1.33 |
| n(0.95,0.90,0.80) | 726,572,413 | 30,23,17 | 8,6,5 | 2,2,2 |

Note that P* is the probability of correct selection for the associated subset selection rule, while the error probability $\gamma$ is controlled at 5 percent level. The identification risk is 0.05 + (1-P*). We can explain the cases described above as follows: for k = 2, if the selected subset contains $\pi_i$ only, we identify $H_i'$, i = 1,2; if it contains $\pi_1$ and $\pi_2$, we identify $H_0$. For k = 3, if the selected subset contains $\pi_i$ only, we identify $H_i'$, i = 1,2,3; if it contains $\pi_1$ and $\pi_2$, $\pi_1$ and $\pi_3$, or $\pi_2$ and $\pi_3$ only, we identify $H_4'$, $H_5'$ or $H_6'$, respectively. Similar discussion applies to the case k = 4.

Now, we discuss the case where $\sigma^2$ is unknown. For any i, the maximum likelihood estimators of $\underline{\tau}_i$ and $\sigma^2$ are:

$$\hat{\underline{\tau}}_i = (C'\Sigma_i^{-1}C)^{-1}C^{-1}\Sigma_i^{-1}\ \underline{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{ik} \end{bmatrix}$$

and

$$\hat{\sigma}^2 = \frac{1}{(k-1)(n-1)}\ \underline{Y}_i'\ [\Sigma_i^{-1} - \Sigma_i^{-1}C(C'\Sigma_i^{-1}C)^{-1}C'\Sigma_i^{-1}]\ \underline{Y}_i .$$

We know that $\hat{\sigma}^2$ and $\hat{\underline{\tau}}_i$ are independent and the distribution f(s) of $s = \frac{\hat{\sigma}}{\sigma}$ is $\sqrt{\chi_p^2(s)}$ with p= (k-1)(n-1).

As before, we define the selection rule as follows:

$$\varphi_i^0(\hat{\underline{\tau}}_i, \hat{\sigma}) = \begin{cases} 1 & \text{if } \ y_{i1} + \ldots + y_{ik} \geq d_1\hat{\sigma}, \\ 0 & \text{otherwise} \end{cases},$$

or

$$\varphi_i^0(\underline{x}, \hat{\sigma}) = \begin{cases} 1 & \text{if } \ \bar{x}_i \geq \frac{1}{k-1} \sum_{j \neq i} \bar{x}_j + \frac{d_1}{k-1}\hat{\sigma} \\ 0 & \text{otherwise} \end{cases}.$$

Conditionally, for an observed value of $\hat{\sigma}$, we can discuss the optimality as before. However, the constant d and n can be determined without any difficulty by (2.8) and (2.9). Since

$$E_{\underline{\tau}}\ \varphi_i^0(\hat{\underline{\tau}}_i, \hat{\sigma}) = \gamma, \qquad \underline{\tau} \in \Omega_0$$

we get

$$\int \Phi\left[-\frac{d_1 s\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}}\right] f(s)ds = \gamma, \tag{2.6}$$

and

$$\inf_{\bar{\Omega}} P(CS|\varphi^0)$$

$$= \int \Phi\left[-\frac{(d_1 s-(k-1)\Delta)\sqrt{n}}{\sqrt{(1-\lambda)k(k-1)}}\right] f(s)ds = P^*. \tag{2.7}$$

This gives

$$t\left[-\frac{d_1\sqrt{n(n-1)}}{\sqrt{(1-\lambda)k}}\ ;\ (k-1)(n-1), 0\right] = \gamma, \tag{2.8}$$

and

$$t\left[-\frac{d_1\sqrt{n(n-1)}}{\sqrt{(1-\lambda)k}}\ ;\ (k-1)(n-1),\ \frac{\Delta\sqrt{n(k-1)}}{\sqrt{(1-\lambda)k}}\right] = P^*, \tag{2.9}$$

where $t(a;\, b,\, c)$ is the percentage point of the noncentral t with b degrees of freedom and the noncentrality parameter c.

Acknowledgment:

The authors wish to thank Professor S. Panchapakesan for a critical reading of this paper and for suggestions to improve the presentation of this paper.

## References

[1] Gupta, S. S. and Huang, D. Y. (1981). Multiple Statistical Decision Theory. Lecture Notes in Statistics (6), Springer-Verlag, New York.

[2] Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures. John Wiley and Sons, New York.

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>Mimeograph Series #81-22 | 2. GOVT ACCESSION NO.<br>A107 921 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Selection Procedures for a Problem in Analysis of Variance | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical<br>6. PERFORMING ORG. REPORT NUMBER<br>Mimeo. Series #81-22 |
| 7. AUTHOR(s)<br>Shanti S. Gupta and Deng-Yuan Huang | | 8. CONTRACT OR GRANT NUMBER(s)<br>ONR N00014-75-C-0455 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Purdue University<br>Department of Statistics<br>West Lafayette, Indiana 47907 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Washington, DC | | 12. REPORT DATE<br>June 1981<br>13. NUMBER OF PAGES<br>16 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified<br>15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION ST... IENT (of ... abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY ...TES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple decision rules, Completely randomized block design, Joint normal distribution, Hypothesis identification, Identification risk, Subset selection, Optimal selection procedures.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

There are many situations in the analysis of variance where an experimenter would like to make comparisons among (and select the "best" set) the treatments. In this paper we study the problem where the data are based on a completely randomized block design. It is shown that the subset selection approach is a useful method to make appropriate "identification" among the hypotheses and the selected subset. We propose an optimal selection procedure which cont·ols the error probabilities when all the parameters (treatments) are equal and which maximizes the infimum of the probability of a correct selection over some preference parameter

space, simultaneously. Some examples are provided to illustrate the optimal subset selection rule and its interpretation in terms of the "identified" hypotheses.