Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

RESTRICTED SUBSET SELECTION FOR NORMAL POPULATIONS WITH UNKNOWN AND UNEQUAL VARIANCES

David W. Sullivan
Lockheed Missiles and Space Co.
P.O. Box 17100
Austin, TX 78760

James R. Wilson
Mechanical Engineering Department
The University of Texas
Austin, TX 78712

ABSTRACT

This paper develops two extensions of the Gupta-Santner restricted subset selection procedure. The (exact) procedure R_E screens a set of k normal populations with unknown and unequal variances using independent random sampling within each alternative population in order to select a final subset of at most m alternatives; in the least favorable configuration of population means, there is the minimal probability P^* that the selected subset includes the population with the largest mean. The simulation-oriented (heuristic) procedure R_S similarly screens a set of k covariance stationary normal processes with unknown and nonidentical covariance structures such that the (correlated) sampling within each alternative process is carried out independently. A rigorous development is given for procedure R_E together with appropriate tables of constants required to apply the rule. The experimental performance of procedure R_S is summarized for a wide variety of stationary autoregressive-moving average processes.

INTRODUCTION

In many experimental situations, the fundamental problem is to select the best of k alternative treatments or populations based on a specified population parameter -- for example, the mean, the variance, a particular quantile, or the percentage of the population falling below a certain cutoff point. In this paper we consider the problem of selecting the population with the largest mean in the case that all populations are normal.

Bechhofer's indifference zone approach [1] to the selection problem was originally formulated for the case of normal populations with a known common variance. This approach allows the experimenter to select a single population so that he has at least the probability P^* of making the correct selection, provided that the population means do not lie in a user-specified indifference zone where it is not practically important to distinguish among alternatives. Also assuming that all populations have a known common variance, Mahamunulu [2] developed an indifference zone procedure to select a subset of fixed size m ($1 \leq m \leq k$) that contains at least c of the t best populations. For the case of normal populations with unknown and unequal variances, Dudewicz and Dalal [3] developed a two-stage version of Bechhofer's procedure; subsequently Koenig and Law [4] developed a similar generalization of Mahamunulu's procedure. For the case of correlated observations within each population, Dudewicz and Zaino [5] proposed a heuristic rule to select the best of k stationary autoregressive processes of order 1. Using spectral analysis, Dickinson [6] extended the Dudewicz-Zaino procedure to handle general covariance stationary processes with unknown and unequal autocovariance functions.

In contrast to the indifference zone approach, Gupta's subset selection approach [7, 8] yields a subcollection of the full set of k alternatives that has a random size and that includes the best population with minimal probability P^* over all possible configurations of the population means. This approach enables the experimenter to screen a large set of alternatives so that the selected subset can be examined more thoroughly in a follow-up study. To reserve adequate resources for the follow-up study, Gupta and Santner [9] devised a restricted subset selection procedure that allows the user to specify an upper bound m on the subset size; see also Santner [10]. Subsequently Santner [11] proposed a restricted subset selection rule with the goal of including at least one of the t best populations in a subset of maximum size m. For k normal populations, all of these procedures require a known common variance. In this paper we develop two extensions of the Gupta-Santner rule: (a) The procedure R_E handles random sampling from normal populations with unknown and unequal variances; and (b) The procedure R_S is designed for use "on the fly" in steady-state simulation, where the output processes of interest are covariance stationary normal processes with unknown and unequal covariance structures. Thus in the context of the restricted subset selection approach, procedures R_E and R_S are respectively analogous to the Dudewicz-Dalal and the Dudewicz-Zaino-Dickinson extensions of the indifference zone approach.

STATEMENT OF THE SELECTION PROBLEM

Suppose that we have a set of k populations $\pi_i \sim N(\mu_i, \sigma_i^2)$, $1 \leq i \leq k$, with unknown $\{\mu_i\}$ and $\{\sigma_i^2\}$. Let $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ denote the ordered means, and let $\pi_{(i)}$ denote the the population with mean $\mu_{[i]}$, $1 \leq i \leq k$. We do not assume any prior information about the correct pairing of the $\{\pi_i\}$ and the $\{\pi_{(i)}\}$. The goal of the rule R_E is to determine the following: (a) the size n_i of the random sample to be taken from π_i, $1 \leq i \leq k$; (b) the statistic $\tilde{X}_i$ to be used as an estimator of the parameter μ_i, $1 \leq i \leq k$; and (c) a screened (selected) subset of candidate populations based on the $\{\tilde{X}_i\}$. The selection rule has three user-specified parameters: (a) the maximal subset size m; (b) the minimal difference δ between the best mean $\mu_{[k]}$ and the next best mean $\mu_{[k-1]}$ that is of practical importance; and (c) the minimal probability P^* of including the best population

$\pi_{(k)}$ in the selected subset whenever there is an "important" difference among the alternatives.

Formally we let Ω denote the underlying configuration space for the vector μ of population means:

$$\Omega \equiv \{\mu = (\mu_1, \ldots, \mu_k) : -\infty < \mu_i < \infty, \text{ all } i\}. \quad (1)$$

We are primarily interested in the preference zone consisting of those configurations with differences that the user considers to be worth detecting:

$$\Omega(\delta) \equiv \{\mu \in \Omega : \mu_{[k-1]} \leq \mu_{[k]} - \delta\}; \quad (2)$$

and in particular we must effectively handle the subspace of least favorable configurations (LFCs):

$$\Omega^*(\delta) \equiv \{\mu \in \Omega(\delta) : \mu_{[1]} = \ldots = \mu_{[k-1]} = \mu_{[k]} - \delta\}. \quad (3)$$

The indifference zone is defined to be the complement $\Omega - \Omega(\delta)$ of the preference zone. If S denotes the size of the selected subset, then the budget constraint on the follow-up study is expressed as

$$1 \leq S \leq m, \quad \text{where} \quad 1 < m < k. \quad (4)$$

(Note that the choice m = 1 reduces to the Dudewicz-Dalal procedure, and the choice m = k reduces to Gupta's original subset selection procedure; hence we only consider the case in which $1 < m < k$.) Finally the probability requirement takes the form

$$\inf_{\mu \in \Omega(\delta)} P_\mu\{CS|R_E\} \geq P^*. \quad (5)$$

THE SELECTION PROCEDURE R_E FOR NORMAL RANDOM SAMPLING

Steps of Procedure R_E

1. Take an initial random sample $\{X_{ij} : 1 \leq j \leq n_0\}$ of size $n_0 \geq 2$ from π_i and calculate the corresponding sample mean and variance:

$$\bar{X}_i = n_0^{-1} \sum_{j=1}^{n_0} X_{ij}, \quad S_i^2 = (n_0-1)^{-1} \sum_{j=1}^{n_0} (X_{ij} - \bar{X}_i)^2. \quad (6)$$

2. Determine $h = h(k,m,P^*,\delta,n_0)$ and $d = d(k,m,P^*,\delta,n_0)$ from Tables 1 and 2 below or compute these quantities as the simultaneous solutions of (29) and (45) below.

3. For population π_i compute the final required sample size

$$n_i = \max\{n_0 + 1, \lceil h^2 S_i^2/d^2 \rceil\}, \text{ where } \lceil r \rceil \equiv \text{smallest integer} \geq r \quad (7)$$

and collect $n_i - n_0$ extra observations from π_i.

4. Compute the second-stage sample means

$$\bar{X}_i' = (n_i - n_0)^{-1} \sum_{j=n_0+1}^{n_i} X_{ij}, \quad 1 \leq i \leq k, \quad (8)$$

and the corresponding weights

$$W_i = (n_0/n_i)\left\{1 + \{[n_i \cdot (hS_i/d)^{-2} - 1] \cdot (n_i - n_0)/n_0\}^{\frac{1}{2}}\right\}, \quad (9)$$

$$W_i' = 1 - W_i, \quad 1 \leq i \leq k, \quad (10)$$

in order to compute the final weighted means

$$\tilde{X}_i = W_i \bar{X}_i + W_i' \bar{X}_i' \quad \text{for} \quad 1 \leq i \leq k. \quad (11)$$

5. Using the ranked weighted means $\tilde{X}_{[1]} \leq \ldots \leq \tilde{X}_{[k]}$, apply the final selection criterion:

$$\text{Select } \pi_i \iff \tilde{X}_i \geq \max\{\tilde{X}_{[k-m+1]}, \tilde{X}_{[k]} - d\}. \quad (12)$$

Probability of Correct Selection for R_E in the LFC

When the population means lie in a least favorable configuration, we must compute the probability of the correct selection event CS, where $CS \equiv \{\pi_{(k)} \text{ is selected}\}$. If i denotes the number of $\tilde{X}$-values that are smaller than the statistic $\tilde{X}_{(k)}$ for the best population $\pi_{(k)}$, then the occurrence of event CS implies that $k-m \leq i \leq k-1$; furthermore for each value of i, there are

$$\binom{k-1}{i} = \frac{(k-1)!}{(k-1-i)! \cdot i!} \quad (13)$$

different ways to select the corresponding populations from among the set $\{\pi_{(1)}, \ldots, \pi_{(k-1)}\}$. For every positive integer g with $i \leq g$, let the sets

$$Q_j(g,i), \quad 1 \leq j \leq C(g,i) \equiv g!/[i! \cdot (g-i)!], \quad (14)$$

represent an enumeration of the distinct subsets of size i that can be taken from the set $\{1, \ldots, g\}$, and let

$$Q_j^c(g,i) \equiv \{1, \ldots, g\} - Q_j(g,i) \quad (15)$$

be the complement of $Q_j(g,i)$ for $1 \leq j \leq C(g,i)$.

The event CS can be decomposed into a finite number of mutually exclusive and exhaustive subevents

$$A_{ij} \equiv \{\tilde{X}_{(u)} < \tilde{X}_{(k)}, \ u \in Q_j(k-1,i); \ \tilde{X}_{(k)} \leq \tilde{X}_{(u)}, \ u \in Q_j^c(k-1,i); \ \tilde{X}_{(k)} + d \geq \tilde{X}_{[k]}\} \quad (16)$$

for $k-m \leq i \leq k-1$ and $1 \leq j \leq C(k-1,i)$, where $\tilde{X}_{(u)}$ is defined to be the statistic associated with population $\pi_{(u)}$. Thus we have:

$$P_\mu\{CS|R_E\} = \sum_{i=k-m}^{k-1} \sum_{j=1}^{C(k-1,i)} P_\mu\{A_{ij}\}, \quad \mu \in \Omega. \quad (17)$$

Now when the event A_{ij} occurs, we have

$$\tilde{X}_{(k)} + d \geq \tilde{X}_{[k]} \geq \tilde{X}_{(u)} \text{ for all } u \in Q_j^c(k-1,i); \quad (18)$$

and it follows that we can express A_{ij} as

$$A_{ij} = \{\tilde{X}_{(u)} < \tilde{X}_{(k)}, \ u \in Q_j(k-1,i); \ \tilde{X}_{(k)} \leq \tilde{X}_{(u)} \leq \tilde{X}_{(k)} + d, \ u \in Q_j^c(k-1,i)\}. \quad (19)$$

To compute $P_\mu\{A_{ij}\}$, we define the variables Y_u and θ_{gu} as follows:

$$\left.\begin{aligned} Y_u &\equiv \{\tilde{X}_{(u)} - \mu_{[u]}\}/(d/h) \\ \theta_{gu} &\equiv \{\mu_{[g]} - \mu_{[u]}\}/(d/h) \end{aligned}\right\} \quad 1 \leq g,u \leq k. \tag{20}$$

Then the relationships in display (16) can be rewritten

$$\tilde{X}_{(u)} < \tilde{X}_{(k)} \iff Y_u < Y_k + \theta_{ku}; \tag{21}$$

$$\tilde{X}_{(k)} \leq \tilde{X}_{(u)} \leq \tilde{X}_{(k)} + d \iff Y_k+\theta_{ku} \leq Y_u \leq Y_k+\theta_{ku}+h. \tag{22}$$

Lemma 4.2 of Mahamunulu [2] can be used to show that the function $\lambda(\mu) \equiv P_\mu\{CS|R_E\}$ is a nonincreasing function of $\mu_{[j]}$, $1 \leq j \leq k-1$ (see Sullivan [12]); thus it follows that

$$\inf_{\mu\in\Omega(\delta)} P_\mu\{CS|R_E\} = \inf_{\mu\in\Omega^{\circ}(\delta)} P_\mu\{CS|R_E\}. \tag{23}$$

Henceforth we assume the LFC, and we suppress the condition $\mu \in \Omega^{\circ}(\delta)$ in all subsequent displays. In this case the θ-values defined by (20) satisfy

$$\theta_{ku} = \{\mu_{[k]} - (\mu_{[k]} - \delta)\}/(d/h) = h\delta/d, \quad 1\leq u\leq k-1. \tag{24}$$

Using the methods of Dudewicz and Dalal [3], Sullivan [12] showed that the Y-variates defined by (20) constitute a random sample of size k from Student's t-distribution with $n_0 - 1$ degrees of freedom:

$$\{Y_u : 1 \leq u \leq k\} \text{ IID} \sim t(n_0 - 1 \text{ d.f.}) \tag{25}$$

The justification of (25) differs from the Dudewicz-Dalal analysis in only one respect: the definitions for the weights $\{W_i, W_i'\}$ and for the $\{Y_u\}$ involve the constant d of procedure R_E rather than the indifference zone width δ.

Applying the law of total probability to each summand of (17) by conditioning on the value of Y_k and combining (21), (22), (24), and (25), we get

$$\inf_{\mu\in\Omega(\delta)} P_\mu\{CS|R_E\} = \sum_{i=k-m}^{k-1} \binom{k-1}{i} \cdot \int_{-\infty}^{\infty} F^i(y+h\delta/d) \cdot [F(y+h\delta/d+h) - F(y+h\delta/d)]^{k-1-i} f(y)dy, \tag{26}$$

where $F(\cdot)$ is the distribution function and $f(\cdot)$ is the density function for Student's t-distribution with n_0-1 degrees of freedom. If we let $b \equiv F(y+h\delta/d)$, $c \equiv F(y+h\delta/d+h)$ and if we reverse the order of integration and summation in (26), then the integrand in (26) has the form

$$\sum_{i=k-m}^{k-1} \binom{k-1}{i} \cdot b^i \cdot (c - b)^{k-1-i}$$

$$= c^{k-1} \cdot \sum_{i=k-m}^{k-1} \binom{k-1}{i} \cdot (b/c)^i \cdot (1 - b/c)^{k-1-i}$$

$$= c^{k-1} \cdot I(b/c;\ k-m,\ m), \quad \text{where} \tag{27}$$

$$I(x;\ p,q) \equiv \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x z^{p-1} \cdot (1-z)^{q-1} dz, \quad 0 \leq x \leq 1 \tag{28}$$

is the incomplete beta function with parameters p and q; see [13], equation (6.6.4). Thus we can express the probability requirement (5) as

$$P^* = \tag{29}$$

$$\int_{-\infty}^{\infty} F^{k-1}(y+h\delta/d+h) \cdot I[F(y+h\delta/d)/F(y+h\delta/d+h);\ k-m,m] \cdot f(y)dy$$

For a given value of d (say, $d = \delta$), we can solve (29) numerically for the constant $h = h(k,m,P^*,\delta,n_0,d)$ that satisfies the probability requirement. To fix a value of d for procedure R_E, we compute the expected subset size as it depends on the values of h and d.

Expected Subset Size for R_E in the LFC

Santner [10] suggested setting d so as to fix the expected value of the subset size S in the least favorable configuration. Let H_u be the indicator function of the event in which the population $\pi_{(u)}$ is selected for the final subset:

$$H_u \equiv \begin{cases} 1 & \text{if } \pi_{(u)} \text{ is selected,} \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

$$\Longrightarrow S = \sum_{u=1}^{k} H_u; \quad \text{and} \tag{31}$$

$$\mu \in \Omega^{\circ}(\delta) \Longrightarrow E_\mu[H_1] = E_\mu[H_u],\ u\leq k-1;\ E_\mu[H_k] = P^*. \tag{32}$$

Thus in the LFC, we have

$$E_\mu[S] = (k - 1)\cdot E_\mu[H_1] + P^*. \tag{33}$$

By analogy to the notation (14) and (15), we make the following definitions for positive integers $g \geq 2$ and $i \leq g$:

$$R_j(g,i) \equiv \text{the } j^{th} \text{ combination of i integers selected from the set } \{2, \ldots, g\}, \text{ and} \tag{34}$$

$$R_j^c(g,i) \equiv \{2, \ldots, g\} - R_j(g,i),\ 1 \leq j \leq C(g,i). \tag{35}$$

To compute $E_\mu[H_1]$, we use an analysis similar to that given for $P_\mu\{CS|R_E\}$. We partition the basic event

$$\{\pi_{(1)} \text{ selected}\} = \{\tilde{X}_{(1)} \geq \tilde{X}_{[k-m+1]};\ \tilde{X}_{(1)} + d \geq \tilde{X}_{[k]}\} \tag{36}$$

into mutually exclusive and exhaustive subevents:

$$B_{ij} \equiv \{\tilde{X}_{(u)} < \tilde{X}_{(1)},\ u\in R_j(k,i);\ \tilde{X}_{(1)} \leq \tilde{X}_{(u)},\ u\in R_j^c(k,i);\ \tilde{X}_{(1)} + d \geq \tilde{X}_{[k]}\}$$

$$= \{\tilde{X}_{(u)} < \tilde{X}_{(1)},\ u \in R_j(k,i);\ \tilde{X}_{(1)} \leq \tilde{X}_{(u)} \leq \tilde{X}_{(1)} + d,\ u \in R_j^c(k,i)\}. \tag{37}$$

As with (21) and (22), we get:

$$\tilde{X}_{(u)} < \tilde{X}_{(1)} \iff Y_u < Y_1 + \theta_{1u}; \tag{38}$$

$$\tilde{X}_{(1)} \leq \tilde{X}_{(u)} \leq \tilde{X}_{(1)}+d \iff Y_1+\theta_{1u} \leq Y_u \leq Y_1+\theta_{1u}+h. \tag{39}$$

Note that in the LFC, we have:

$$\theta_{1u} = 0,\ 1 \leq u \leq k-1; \quad \theta_{1k} = -h\delta/d. \tag{40}$$

There are two conditions which allow us to further partition B_{ij}. We can first consider the case in which the Y-value associated with the best alternative $\pi_{(k)}$ is greater than the Y-value associated with the worst alternative $\pi_{(1)}$. In this case, we combine (38), (39), and (40) to define

$$\begin{aligned} B_{ij}^{+} \equiv \{&Y_u < Y_1,\ u \in R_j(k-1,i); \\ &Y_1 \leq Y_u \leq Y_1 + h,\ u \in R_j^c(k-1,i); \\ &Y_1 - h\delta/d \leq Y_k \leq Y_1 - h\delta/d + h\}. \end{aligned} \tag{41}$$

When $Y_1 \leq Y_k$, there are $C(k-2,i)$ ways to select the i alternatives with $Y_u < Y_1$.

Next we consider the case in which $Y_1 > Y_k$. There are $C(k-2,i-1)$ different events of the form

$$\begin{aligned} B_{ij}^{-} \equiv \{&Y_u < Y_1,\ u \in R_j(k-1,i-1);\ Y_k < Y_1 - h\delta/d; \\ &Y_1 \leq Y_u \leq Y_1 + h,\ u \in R_j^c(k-1,i-1)\}, \end{aligned} \tag{42}$$

where again we make use of (38), (39), and (40). To compute $P_\mu\{B_{ij}^{+}\}$ and $P_\mu\{B_{ij}^{-}\}$, we apply the law of total probability to (41) and (42) by conditioning on the value of Y_1; then combining all of the terms composing the event $\{\pi_{(1)}$ is selected$\}$, we get

$$\begin{aligned} E_\mu[H_1] &= \sum_{i=k-m}^{k-2} \binom{k-2}{i} \cdot P_\mu\{B_{i1}^{+}\} + \sum_{i=k-m}^{k-1} \binom{k-2}{i-1} \cdot P_\mu\{B_{i1}^{-}\} \\ &= \int_{-\infty}^{\infty} [F(y-h\delta/d+h) - F(y-h\delta/d)] \cdot F^{k-2}(y+h) \cdot \\ &\qquad I[F(y)/F(y+h);\ k-m,m-1] \cdot f(y)dy \\ &+ \int_{-\infty}^{\infty} F(y-h\delta/d) \cdot F^{k-2}(y+h) \cdot I[F(y)/F(y+h);\ k-m-1,m] f(y)dy \end{aligned} \tag{43}$$

For the procedure R_E we take the expected subset size to be the midrange of the acceptable sizes specified by the user:

$$E_\mu[S] = (m+1)/2; \tag{44}$$

and combining (33), (43), and (44), we finally obtain:

$$\begin{aligned} &\int_{-\infty}^{\infty} [F(y-h\delta/d+h) - F(y-h\delta/d)] \cdot F^{k-2}(y+h) \cdot \\ &\qquad I[F(y)/F(y+h);\ k-m,m-1] \cdot f(y)dy \\ &+ \int_{-\infty}^{\infty} F(y-h\delta/d) \cdot F^{k-2}(y+h) \cdot I[F(y)/F(y+h);\ k-m-1,m] \cdot f(y)dy \\ &= [(m+1)/2 - P^*]/(k-1). \end{aligned} \tag{45}$$

An iterative numerical method for computing the solution $\{h, d\}$ of equations (29) and (45) is described in the next subsection.

Tables for Procedure R_E

Since δ only appears as a divisor of d in (29) and (45) Tables 1 and 2 show how the constants h and $d' \equiv d/\delta$ depend on k, m, P^*, and n_0. The required values of d can be computed from $d = d' \cdot \delta$. The table values were generated by an iterative algorithm producing successively closer approximations to the simultaneous solutions of equations (29) and (45). The algorithm is as follows:

1. Initialize the current trial solution

$$\hat{h}_1 = \hat{h}_0, \qquad \hat{d}_1' = \hat{d}_0', \tag{46}$$

where $\hat{h}_0$ and $\hat{d}_0'$ are initial values selected so as to accelerate the convergence of the algorithm. With no other information available, acceptable starting values are $\hat{h}_0 = 1.0$, $\hat{d}_0' = 1.0$. Better starting values are given by the solutions of relevant previous cases:

$$\hat{h}_0(k,m,P^*,n_0) = h(k-1,m,P^*,n_0), \tag{47}$$

$$\hat{d}_0'(k,m,P^*,n_0) = d'(k,m,P^*,n_0-\Delta n_0). \tag{48}$$

Such a bootstrapping scheme was used in generating Tables 1 and 2.

2. Solve equation (29) for a new h-value $\hat{h}_2$ given the current d-value $\hat{d}_1'$. We used the IMSL routine ZSCNT [14] for solving a nonlinear equation by the secant method, and we specified a maximum absolute error of 5.0×10^{-6} in the estimated solution. To evaluate the integral in (29), we used the IMSL numerical quadrature routine DCADRE [14] with maximum absolute error $1.\times10^{-6}$ and maximum relative error 5.0×10^{-4}.

3. Solve equation (45) for a new d-value $\hat{d}_2$ given the new h-value $\hat{h}_2$. We used the IMSL routines ZSCNT and DCADRE in the same way as for step 2.

4. Compute the increments

$$\Delta\hat{h} = \hat{h}_2 - \hat{h}_1, \qquad \Delta\hat{d}' = \hat{d}_2' - \hat{d}_1'; \tag{49}$$

and test the termination condition

$$|\Delta\hat{h}| < \varepsilon \quad \text{and} \quad |\Delta\hat{d}'| < \varepsilon'. \tag{50}$$

If (50) is satisfied, deliver the final solutions $h \leftarrow \hat{h}_2'$ and $d \leftarrow \hat{d}_2'$ and stop; otherwise go to step 5. We used the tolerances $\varepsilon = \varepsilon' = 10^{-4}$ in computing Tables 1 and 2.

5. Update the current h- and d-values

$$\hat{h}_1 \leftarrow \hat{h}_2, \qquad \hat{d}_1' \leftarrow \hat{d}_2' \tag{51}$$

and go to step 2.

To generate additional h- and d-values for parameter sets $\{m, k, P^*, n_0\}$ not appearing in this paper, we

Table 1: Values of h for Selection Procedure R_E

k	m	P* = 0.90 n_0=10	n_0=20	n_0=30	P* = 0.95 n_0=10	n_0=20	n_0=30	P* = 0.99 n_0=10	n_0=20	n_0=30
3	2	1.173	1.098	1.077	1.483	1.369	1.337	2.333	2.089	2.026
4	2	1.175	1.091	1.067	1.517	1.390	1.356	2.423	2.162	2.096
	3	1.354	1.268	1.244	1.612	1.486	1.452	2.302	2.041	1.973
5	2	1.179	1.088	1.063	1.535	1.402	1.367	2.461	2.193	2.125
	3	1.326	1.230	1.204	1.605	1.467	1.429	2.358	2.075	2.003
	4	1.499	1.403	1.375	1.735	1.599	1.562	2.359	2.087	2.017
6	2	1.183	1.087	1.061	1.547	1.410	1.374	2.483	2.210	2.141
	3	1.313	1.210	1.182	1.606	1.458	1.419	2.391	2.098	2.024
	4	1.458	1.353	1.324	1.710	1.562	1.522	2.384	2.088	2.013
	5	1.616	1.510	1.480	1.839	1.694	1.654	2.428	2.146	2.075
7	2	1.185	1.086	1.059	1.555	1.415	1.379	2.497	2.221	2.151
	3	1.306	1.197	1.167	1.608	1.454	1.413	2.413	2.115	2.040
	4	1.436	1.323	1.292	1.698	1.541	1.499	2.404	2.095	2.017
	5	1.569	1.456	1.424	1.805	1.648	1.606	2.432	2.128	2.052
8	2	1.187	1.085	1.058	1.561	1.419	1.383	2.507	2.227	2.158
	3	1.302	1.188	1.157	1.610	1.451	1.410	2.428	2.126	2.051
	4	1.422	1.303	1.270	1.692	1.527	1.484	2.420	2.102	2.023
	5	1.541	1.421	1.387	1.785	1.620	1.576	2.442	2.122	2.043
9	2	1.189	1.085	1.058	1.566	1.423	1.386	2.514	2.232	2.163
	3	1.300	1.182	1.150	1.612	1.450	1.407	2.440	2.135	2.060
	4	1.413	1.288	1.254	1.689	1.518	1.473	2.432	2.109	2.028
	5	1.523	1.396	1.361	1.774	1.601	1.555	2.451	2.122	2.040
10	2	1.191	1.085	1.058	1.570	1.425	1.389	2.520	2.237	2.167
	3	1.298	1.177	1.144	1.614	1.449	1.406	2.449	2.142	2.067
	4	1.406	1.277	1.242	1.687	1.512	1.465	2.443	2.115	2.034
	5	1.510	1.377	1.341	1.766	1.587	1.540	2.460	2.123	2.039
15	2	1.198	1.085	1.057	1.582	1.433	1.397	2.538	2.248	2.178
	3	1.297	1.162	1.127	1.623	1.448	1.403	2.477	2.164	2.090
	4	1.392	1.246	1.207	1.687	1.495	1.446	2.475	2.136	2.054
	5	1.480	1.327	1.286	1.753	1.552	1.500	2.490	2.134	2.047
20	2	1.203	1.085	1.057	1.590	1.437	1.402	2.549	2.254	2.185
	3	1.298	1.156	1.119	1.630	1.448	1.404	2.492	2.176	2.102
	4	1.389	1.231	1.190	1.691	1.489	1.438	2.493	2.149	2.066
	5	1.470	1.303	1.259	1.752	1.538	1.483	2.509	2.143	2.055
30	2	1.210	1.086	1.058	1.601	1.442	1.407	2.562	2.260	2.190
	3	1.303	1.149	1.111	1.641	1.449	1.405	2.511	2.188	2.114
	4	1.389	1.217	1.173	1.699	1.484	1.432	2.515	2.163	2.082
	5	1.465	1.280	1.232	1.756	1.525	1.467	2.533	2.156	2.069
40	2	1.216	1.086	1.058	1.607	1.445	1.409	2.570	2.263	2.194
	3	1.308	1.146	1.107	1.649	1.451	1.406	2.522	2.195	2.123
	4	1.392	1.209	1.163	1.707	1.482	1.429	2.530	2.171	2.092
	5	1.466	1.268	1.218	1.762	1.519	1.460	2.549	2.164	2.079
50	2	1.220	1.087	1.058	1.613	1.447	1.411	2.577	2.265	2.196
	3	1.313	1.144	1.105	1.655	1.452	1.408	2.531	2.200	2.127
	4	1.396	1.205	1.158	1.713	1.481	1.428	2.540	2.177	2.099
	5	1.469	1.261	1.209	1.768	1.516	1.456	2.560	2.170	2.085

Table 2: Values of d/δ for Selection Procedure R_E

k	m	P* = 0.90 n_0=10	n_0=20	n_0=30	P* = 0.95 n_0=10	n_0=20	n_0=30	P* = 0.99 n_0=10	n_0=20	n_0=30
3	2	.743	.736	.734	.740	.730	.728	.798	.786	.782
4	2	.609	.598	.595	.641	.629	.625	.729	.714	.710
	3	.898	.896	.895	.849	.844	.842	.851	.841	.838
5	2	.547	.535	.532	.591	.578	.575	.690	.675	.670
	3	.746	.737	.734	.738	.726	.723	.779	.763	.758
	4	1.004	1.006	1.007	.925	.924	.924	.891	.885	.884
6	2	.509	.497	.494	.559	.546	.543	.664	.648	.644
	3	.667	.655	.651	.677	.663	.659	.737	.718	.713
	4	.851	.845	.843	.814	.806	.804	.820	.806	.803
	5	1.082	1.088	1.090	.980	.984	.985	.921	.919	.919
7	2	.483	.471	.468	.536	.524	.521	.644	.628	.625
	3	.617	.603	.599	.637	.622	.617	.708	.688	.683
	4	.764	.754	.751	.748	.736	.733	.776	.758	.754
	5	.933	.931	.930	.873	.869	.868	.853	.843	.840
8	2	.463	.451	.448	.519	.507	.504	.628	.613	.609
	3	.583	.568	.563	.609	.592	.588	.686	.666	.660
	4	.707	.694	.691	.704	.689	.685	.746	.725	.720
	5	.843	.836	.834	.806	.797	.795	.809	.794	.790
9	2	.448	.436	.433	.505	.493	.491	.616	.601	.597
	3	.557	.541	.536	.587	.570	.565	.669	.648	.643
	4	.667	.652	.647	.672	.655	.650	.723	.701	.695
	5	.783	.772	.769	.760	.747	.743	.778	.759	.754
10	2	.436	.424	.421	.494	.482	.480	.605	.590	.587
	3	.536	.520	.515	.570	.552	.548	.655	.634	.629
	4	.636	.620	.615	.647	.629	.624	.704	.682	.675
	5	.738	.725	.721	.725	.710	.705	.754	.733	.727
15	2	.396	.386	.383	.455	.446	.445	.569	.556	.553
	3	.475	.458	.453	.516	.499	.494	.609	.589	.585
	4	.548	.528	.523	.574	.553	.547	.648	.623	.617
	5	.618	.598	.593	.629	.607	.601	.684	.658	.651
20	2	.373	.364	.362	.433	.425	.424	.547	.535	.533
	3	.442	.425	.420	.486	.469	.465	.583	.564	.560
	4	.505	.483	.477	.536	.514	.508	.616	.592	.586
	5	.562	.540	.533	.581	.558	.551	.646	.620	.613
30	2	.347	.339	.338	.406	.400	.400	.519	.509	.508
	3	.406	.389	.385	.452	.436	.433	.550	.533	.530
	4	.458	.436	.430	.493	.472	.466	.579	.556	.551
	5	.504	.479	.472	.531	.506	.499	.604	.578	.571
40	2	.331	.324	.323	.389	.385	.385	.501	.493	.493
	3	.385	.368	.364	.431	.417	.414	.530	.515	.512
	4	.431	.409	.403	.469	.448	.443	.556	.535	.530
	5	.472	.447	.440	.502	.477	.470	.580	.553	.548
50	2	.319	.313	.313	.377	.374	.375	.489	.482	.482
	3	.370	.354	.351	.417	.403	.401	.516	.501	.499
	4	.414	.391	.386	.452	.431	.427	.540	.520	.516
	5	.452	.425	.418	.483	.458	.451	.562	.537	.532

recommend using Tables 1 and 2 to set appropriate start-values as prescribed by (47) and (48).

SELECTION PROCEDURE R_S FOR STATIONARY NORMAL PROCESSES

Motivation

If $\pi_i = \{X_{ij} : j \geq 1\}$ is the output process generated by the ith in a set of k alternative steady-state simulation models, then it is reasonable to assume that π_i is covariance stationary with mean μ_i, variance σ_i^2, lag-ℓ covariance

$$\gamma_i(\ell) \equiv E[(X_{ij} - \mu_i)\cdot(X_{i,j+\ell} - \mu_i)], \tag{52}$$

and the summability property

$$\sum_{\ell=-\infty}^{\infty} |\gamma_i(\ell)| < \infty \tag{53}$$

so that we can also define the associated covariance parameter

$$\gamma_i \equiv \sum_{\ell=-\infty}^{\infty} \gamma_i(\ell) \tag{54}$$

and the power spectrum

$$p_i(\omega) \equiv \sum_{\ell=-\infty}^{\infty} \gamma_i(\ell)\cdot\exp[-2\pi\cdot\ell\cdot\omega\cdot\sqrt{(-1)}], \quad -\tfrac{1}{2} \leq \omega \leq \tfrac{1}{2}. \tag{55}$$

Now if each alternative π_i is an autoregressive process of order 1

$$\left.\begin{aligned} X_{ij} &= \mu_i + \phi_i(X_{i,j-1} - \mu_i) + Z_{ij}, \quad j \geq 1; \\ |\phi_i| &< 1 \quad \text{and} \quad \{Z_{ij} : j \geq 1\} \text{ IID} \sim N(0, \sigma_{Zi}^2) \end{aligned}\right\} \tag{56}$$

then the Dudewicz-Zaino procedure A [5] for selecting the process with the largest mean consists essentially of the following steps: (a) Compute the sample size n_i required by the first stage of the Dudewicz-Dalal procedure P_E [3] together with the first-stage estimator $\hat{\phi}_i$ of the autoregressive parameter ϕ_i; (b) Inflate the required sample size according to the formula

$$n_i^* = n_i\cdot(1 + \hat{\phi}_i)/(1 - \hat{\phi}_i) \tag{57}$$

and take $n_i^* - n_0$ additional observations on π_i; and finally (c) Select the process with the largest (unweighted) sample mean

$$\bar{X}_i^* = (n_i^*)^{-1} \sum_{j=1}^{n_i^*} X_{ij}, \qquad 1 \leq i \leq k. \tag{58}$$

The significance of the inflation factor $(1+\hat{\phi}_i)/(1-\hat{\phi}_i)$ is that it captures the effect of the covariance structure of an AR(1) process as follows:

$$(1 + \phi_i)/(1 - \phi_i) = \gamma_i/\gamma_i(0) = p_i(0)/\sigma^2; \tag{59}$$

thus the required sample size n_i for independent sampling under procedure P_E must be adjusted by the number of correlated observations that are "equivalent" to one independent observation in order to yield the final sample size n_i^* for procedure A.

From the foregoing discussion it follows that an extension of the restricted subset selection procedure R_E to handle general covariance stationary processes requires a first-stage estimator $\tilde{\gamma}_i$ for $\gamma_i = p_i(0)$ to replace S_i^2 in equation (7) with the following properties: (a) $\tilde{\gamma}_i$ should be an appropriately scaled chi-square variate with degrees of freedom ν_0 common to all populations -- at least asymptotically as the first-stage sample size becomes sufficiently large:

$$\tilde{\gamma}_i \xrightarrow[n_0\to\infty]{D} \gamma_i\cdot\chi^2(\nu_0 \text{ d.f.})/\nu_0 \quad \text{for} \quad 1 \leq i \leq k; \tag{60}$$

and (b) $\tilde{\gamma}_i$ and the final statistic $\bar{X}_i^*$ should be asymptotically independent with

$$(\bar{X}_i^* - \mu_i)/(d/h) \xrightarrow[n_0\to\infty,\ \delta\to 0]{D} t(\nu_0 \text{ d.f.}). \tag{61}$$

These considerations motivate the heuristic procedure R_S described below.

Steps of Procedure R_S

1. Take an initial sample $\{X_{ij} : 1 \leq j \leq n_0\}$ of size $n_0 \geq 120$ from π_i and compute the Heidelberger-Welch estimator $\tilde{\gamma}_i$ [15] of the spectrum at zero frequency $p_i(0)$.

   a. Compute the periodogram

$$I_i(\omega_u) = n_0^{-1} \left| \sum_{j=1}^{n_0} X_{ij}\cdot\exp[-2\pi\cdot(j-1)\cdot\omega_u\cdot\sqrt{(-1)}] \right|^2 \tag{62}$$

   at the frequencies $\omega_u \equiv u/n_0$, $0 < u < n_0$.

   b. Average adjacent periodogram values to yield "normalized" variates, take logarithms to stabilize the variance, and adjust for the bias introduced by the logarithmic transformation:

$$Y_{iu} = \ln\{\tfrac{1}{2}[I_i((2u-1)/n_0) + I_i(2u/n_0)]\} + \ln(2) - \psi(2) \qquad 1 \leq u \leq 50, \tag{63}$$

   where $\psi(z) \equiv \Gamma'(z)/\Gamma(z)$ is the digamma function [13, Table 6.1].

   c. Using the independent variables

$$G_{iuv} = [(4u-1)/(2n_0)]^v, \quad 0 \leq v \leq 2, \tag{64}$$

   fit a quadratic polynomial to the first 25 values of the series $\{Y_{iu} : 1 \leq u \leq 50\}$ by the method of ordinary least squares:

$$Y_{iu} = \sum_{v=0}^{2} \beta_{iv}G_{iuv} + \varepsilon_{iu}, \quad 1 \leq u \leq 25. \tag{65}$$

   Let $\underset{\sim}{G}_i$ denote the matrix $\|G_{iuv}\|$ having the element G_{iuv} in row u and column v, and let $\hat{\beta}_{i0}$ denote the

least-squares estimate of the intercept in (65).

d. Compute the Heidelberger-Welch estimator

$$\hat{\gamma}_i = \exp\{\hat{\beta}_{i0} - \tfrac{1}{2}\psi'(2)\cdot[(\underset{\sim}{G}_i'\underset{\sim}{G}_i)^{-1}]_{11}\}. \qquad (66)$$

Note that as $n_0 \to \infty$, $\hat{\gamma}_i$ is approximately distributed as $\gamma_i \cdot \chi^2(\nu_i)/\nu_i$, where the "effective degrees of freedom" ν_i is given by

$$\nu_i = 2/(\exp\{\psi'(2)\cdot[(\underset{\sim}{G}_i'\underset{\sim}{G}_i)^{-1}]_{11}\} - 1); \qquad (67)$$

moreover, $\hat{\gamma}_i$ is independent of the sample mean of the original process $\{X_{ij} : j \geq 1\}$.

2. Take ν_i as the degrees of freedom for Student's t-distribution in equations (29) and (45); then as in procedure R_E, solve these equations simultaneously for h and d.

3. Calculate the final required sample size

$$n_i^* = \max\{n_0, \lceil h^2\hat{\gamma}_i/d^2\rceil\} \qquad (68)$$

and collect $n_i^* - n_0$ additional observations from π_i.

4. Compute the final (unweighted) sample means according to (58).

5. Using the ranked sample means $\bar{X}^*_{[1]} \leq \ldots \leq \bar{X}^*_{[k]}$, apply the selection criterion:

$$\text{Select } \pi_i \iff \bar{X}_i^* \geq \max\{\bar{X}^*_{[k-m+1]}, \bar{X}^*_{[k]} - d\}. \qquad (69)$$

Experimental Performance of Procedure R_S

The selection rule R_S was tested using various sets of time series generated from autoregressive-moving average (ARMA) processes of the form

$$X_{ij} = \mu_i + \sum_{\ell=1}^{p} \phi_{i\ell}(X_{i,j-\ell}-\mu_i) + Z_{ij} - \sum_{\ell=1}^{q} \theta_{i\ell}Z_{i,j-\ell} \qquad (70)$$

with $\{Z_{ij} : j \geq 1\}$ IID $\sim N(0, \sigma_{Zi}^2)$ for $1 \leq i \leq k$,

such that each process π_i is stationary and invertible [16]; moreover, the process means $\{\mu_i\}$ lie in the LFC. For each case, δ was selected as a fraction of $\max\{\sqrt{\gamma_i} : 1 \leq i \leq k\}$. The experimental protocol for evaluating the performance of procedure R_S consists of the following steps:

1. Initialize π_i with $X_{ij} = \mu_i$, $-(p-1) \leq j \leq 0$, and $Z_{ij} = 0$, $-(q-1) \leq j \leq 0$; and "warm up" the process by generating (and discarding) an initial series of 100 observations. For the white noise process $\{Z_{ij} : j \geq 1\}$ driving π_i, the IMSL normal deviate generator GGNML [14] is used. The purpose of this step is to eliminate transient effects from each process.

2. Generate the first-stage time series $\{X_{ij} : 1 \leq j \leq n_0\}$ of length n_0 for π_i, $1 \leq i \leq k$; apply rule R_S; and record the occurrence or nonoccurrence of event CS.

3. Independently reseed the IMSL random number generator GGUBS [14] and repeat steps 1 and 2 until a complete meta-experiment consisting of 50 independent replications of procedure R_S has been performed.

Table 3 presents a condensed description of the layouts of the various ARMA processes that we used, and Table 4 summarizes the results of each meta-experiment.

Table 3: Summary of Layouts of ARMA Processes Used

Layout	k	$\max\{\gamma_i\}$	$\min\{\gamma_i\}$
A	40	2.78	0.36
B	10	100.0	0.00277
C	40	100.0	0.00277
D	25	100.0	0.00277

Table 4: Experimental Performance of Procedure R_S

Meta-expt.	Layout	k	m	P*	δ	n_0	%CS
1	A	40	5	0.95	0.23	90	0.90
2	A	40	5	0.98	0.23	90	0.98
3	A	40	2	0.95	0.23	90	0.94
4	A	40	2	0.95	0.23	120	0.96
5	B	10	2	0.95	1.0	30	0.96
6	B	10	2	0.95	1.0	60	0.98
7	B	10	2	0.95	1.0	90	1.0
8	B	10	2	0.95	1.0	120	0.98
9	B	10	2	0.99	1.0	30	0.94
10	B	10	2	0.99	1.0	60	0.98
11	B	10	2	0.99	1.0	90	1.00
12	B	10	2	0.95	0.5	90	0.94
13	C	25	3	0.95	0.5	90	0.96
14	C	25	2	0.98	0.5	90	0.96
15	C	25	3	0.95	0.25	90	0.80
16	C	25	3	0.95	0.25	120	0.98
17	C	25	3	0.95	1.0	60	0.94
18	C	25	3	0.95	1.0	90	0.96
19	D	40	5	0.95	1.0	30	0.86
20	D	40	5	0.95	1.0	60	0.92
21	D	40	5	0.95	1.0	90	0.98
22	D	40	5	0.95	1.0	120	0.98

The results in Table 4 show that the probability requirement P* is closely matched by the observed relative frequency of correct selection (%CS), although degraded performance is observed for small values of n_0. This is attributable to the poor behavior of the Heidelberger-Welch estimator $\hat{\gamma}_i$ of the spectrum at zero frequency when fewer than 25 points are available for fitting a quadratic polynomial to the log-periodogram (63). Since an initial sample of size n_0 yields only $n_0/4$ points of the log-periodogram, we conclude that in the first stage of procedure R_S the experimenter should take $n_0 \geq 120$.

In a meta-experiment consisting of r = 50 independent replications, the standard error of the estimator %CS is

$$SE[\ \%CS\ ] = [P^*\cdot(1 - P^*)/r]^{\frac{1}{2}}; \qquad (71)$$

and when $n_0 = 120$, we observe that %CS falls within 2 standard errors of the nominal correct-selection probability P* in each layout of ARMA processes shown in Table 4. These results strongly suggest that procedure R_S is a viable solution to the restricted subset selection problem in the context of steady-state simulation. We are continuing our development and analysis of this procedure.

REFERENCES

[1] Bechhofer, R. E., "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," Annals of Mathematical Statisitcs, 25, 16-39, 1954.

[2] Mahamunulu, D. M., "Some Fixed-Sample Ranking and Selection Problems," Annals of Mathematical Statistics, 38, 1079-1091, 1967.

[3] Dudewicz, E. J. and Dalal, S. R., "Allocation of Observations in Ranking and Selection with Unequal Variances," Sankhya, B37, 28-78, 1975.

[4] Koenig, L. W. and Law, A. M., "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, with Applications to Simulation," Technical Report No. 82-9, Department of Management Information Systems, University of Arizona, Tucson, Arizona, 1982.

[5] Dudewicz, E. J., and Zaino, N. A., "Allowance for Correlation in Setting Simulation Run-Length via Ranking-and-Selection Procedures," TIMS Studies in the Management Sciences, Volume 7: Algorithmic Methods in Probability, North-Holland Publishing Company, Amsterdam, 1977, 51-61.

[6] Dickinson, R. T., "A Multiple Ranking Procedure Adapted to Discrete-Event Simulation," Technical Report, Department of Mechanical Engineering, University of Texas, Austin, Texas, 1983.

[7] Gupta, S. S., "On a Decision Rule for a Problem in Ranking Means," Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina, 1956.

[8] Gupta, S. S., "On Some Multiple Decision (Ranking and Selection) Rules," Technometrics, 7, 225-245, 1965.

[9] Gupta, S. S., and Santner, T. J., "On Selection and Ranking Procedures -- A Restricted Subset Selection Rule," Proceedings of the 39th Session of the International Statistical Institute, Vienna, 1973, 478-486.

[10] Santner, T. J., "A Restricted Subset Selection Approach to Ranking and Selection Problems," Annals of Statistics, 3, 334-349, 1975.

[11] Santner, T. J., "A Generalized Goal in Restricted Subset Selection Theory," Sankhya, B38, 129-143, 1976.

[12] Sullivan, D. W., "A Restricted Subset Selection Procedure Adapted to Discrete-Event Simulation," Technical Report, Department of Mechanical Engineering, University of Texas, Austin, Texas, 1984.

[13] Abramowitz, M., and Stegun, I. A., Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series No. 55, Washington, D. C., 1972.

[14] International Mathematical and Statistical Library, IMSL Library Reference Manual, 8th Edition, IMSL, Inc., Houston, Texas, 1980.

[15] Heidelberger, P. and Welch, P. D., "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," Communications of the ACM, 24, 233-245, 1981.

[16] Box, G. E. P., and Jenkins, G. M., Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco, California, 1976.