Chapter 5

SUBSET SELECTION OF THE T BEST POPULATIONS

5.1 Introduction

In Chapter 4 we discussed the problem of selecting the t "good" populations by choosing those t populations with the largest observed location parameters (and smallest observed scale parameters). Lower bounds were obtained on the difference between the minimum selected and maximum non-selected parameters, and upper and lower bounds were obtained on all parameters. As indicated in that chapter, improved bounds may be obtained if attention is restricted to a particular selection problem. This is investigated in the present chapter.

Suppose that we wish to select a (random size) subset containing the t good populations, with a prespecified probability of correct selection (PCS). This selection goal has been previously investigated by Carroll, Gupta and Huang (1975) for the location and scale parameter cases, and by Gupta and Sobel (1962) for the particular case of normal variances. Carroll, Gupta and Huang's lower bound on the PCS for the location parameter case is discussed in Section 5.2 and is compared with an alternative approach taken by Bofinger and Mengersen (1986). It is the latter approach that is studied in detail in this chapter. Although the new results still fail to provide an exact solution, the corresponding lower bound on the PCS is

shown to be less conservative than that of Carroll, Gupta and Huang.

Special results are obtained for the case of normal means with common unknown variance and the relevant tables are presented. Bofinger and Mengersen investigated the conservativeness of their bound on the PCS for this particular case using simulation and, based on these results, conjectured a new lower bound for sufficiently large PCS.

Carroll, Gupta and Huang's results for the scale parameter case are compared in Section 5.3 with those obtained by Bofinger and Mengersen. Particular reference is made to the problem of selecting the t normal populations with the smallest variances and the necessary tables are produced. As for the location parameter case, Bofinger and Mengersen's results are shown to be superior to those of Carroll, Gupta and Huang's. Gupta and Sobel's approach is also discussed; their conjecture (Conjecture 2 in their paper) is analagous to that made by Bofinger and Mengersen for the location parameter case and, if true, would result in a less conservative bound on the PCS than those of both Carroll, Gupta and Huang, and Bofinger and Mengersen.

In Section 5.4, a related selection problem is considered. This involves selection of *only* good populations (rather than all good populations) and was investigated by Bofinger and Mengersen. An example illustrating the results of Section 5.2 (for the location parameter case) is given in Section 5.5.

5.2 Selection of all t good populations - location parameter case

Consider k random variables X_i, $i = 1, .., k$, from populations π_i with continuous distribution functions $F(x - \theta_i)$. As usual, we order the θ_i and X_i values as follows:

$$\theta_{\rho(1)} \leq \theta_{\rho(2)} \leq \ldots \leq \theta_{\rho(k-t)} < \theta_{\rho(k-t+1)} \leq \ldots \leq \theta_{\rho(k)}$$

$$X_{R(1)} < X_{R(2)} < \ldots < X_{R(k)}$$

with $\rho(\cdot)$ and $R(\cdot)$ parametric and random permutation functions respectively.

5.2.1 A previous approach

The goal considered by Carroll, Gupta and Huang (1975) is to select a subset of populations, based on the X_i values, of size at least t, containing the t good populations, that is, those π_i corresponding to the t largest θ_i values. The selection goal, then, is:

Goal G1: To select populations $\pi_i \quad \forall i \in \gamma_t$, where

$$\gamma_t = \{\rho(k-t+1), \rho(k-t+2), .., \rho(k)\} \ . \tag{5.1}$$

The selection procedure R_3 considered by Carroll, Gupta and Huang may be written as:

Procedure P1: Select a subset of populations $\pi_i \quad \forall i \in G(d)$, where

$$G(d) = \{i : X_i > X_{R(k-t+1)} - d\} \tag{5.2}$$

with d a positive constant.

A correct selection, CS (referred to by Carroll, Gupta and Huang as a correct decision, CD) is defined to be the event

$$CS = \{\gamma_t \subseteq G(d)\} \ , \tag{5.3}$$

that is, that all t populations in γ_t are selected.

The problem, then, is to determine the smallest value of d such that the PCS is at least a specified value, so that

$$PCS = P_{\underline{\theta}}\{CS \mid P1\} = P_{\underline{\theta}}\{\gamma_t \subseteq G(d)\} \geq P^* \ . \tag{5.4}$$

Lower bound on the PCS

Under their procedure R_3 (given by P1 here), Carroll, Gupta and Huang were unable to find an explicit expression for the LFC of means (at which the PCS attains its infimum), so instead they derived a lower bound on the PCS by considering the event

$$E = \{X_i - \theta_i > X_j - \theta_j - d \quad \forall i \in \gamma_t, j \notin \gamma_t\} \tag{5.5}$$

with γ_t defined in (5.1).

Since the event that these bounds are correct is a subset of the event CS of correct selection, the probability of the event E is a lower bound for the PCS under procedure P1. As discussed in Chapter 4, E has probability given by

$$P_{k,t}(d) = t \int (F(x+d))^{k-t}(1-F(x))^{t-1} dF(x) \tag{5.6}$$

and hence Carroll, Gupta and Huang showed that

$$\inf_{\theta \in \Omega} P_\theta\{CS \mid P1\} \geq P_{k,t}(d) \ . \tag{5.7}$$

A Monte Carlo study conducted by Carroll, Gupta and Huang for the case of normal means with equal known variance, however, showed that this lower bound is quite conservative, due to the use of the "catch-all" event E.

Upper bound on the PCS

An upper bound on the infimum of the PCS was also given by Carroll, Gupta and Huang in their Lemma 3.2. Under the configuration

$$\theta_i = \theta_j + \delta \quad \forall i \in \gamma_t, j \notin \gamma_t \ ,$$

the authors showed that

$$\begin{aligned} P\{CD \mid R_3\} &= t\Sigma_{j=0}^{t-1} \binom{t-1}{j} \binom{k-t}{t-j-1} A_1(k,t,j) \\ &+ (k-t)\Sigma_{j=0}^{t-1} \binom{t}{j} \binom{k-t-1}{t-j-1} A_2(k,t,j) \end{aligned} \tag{5.8}$$

where

$$\begin{aligned} A_1(k,t,j) &= \int (1-F(x))^j (1-F(x+\delta))^{t-j-1} (F(x)-F(x-d))^{t-j-1} \\ &\quad (F(x+\delta))^{k-2t+j+1} dF(x) \\ A_2(k,t,j) &= \int (1-F(x-\delta))^j (F(x-\delta)-F(x-\delta-d))^{t-j} \\ &\quad (1-F(x))^{t-j-1} (F(x))^{k-2t+j} dF(x) \ . \end{aligned}$$

Critical values satisfying (5.8) for given P^* were tabulated by Carroll, Gupta and Huang (in their Table III) for the normal means case. This upper bound on the infimum of the PCS was improved by Bofinger and Mengersen (1986), as discussed later in this chapter.

5.2.2 A new approach

Lower bound on the PCS

The problem of selecting the t best populations was considered using a different approach by Bofinger and Mengersen (1986). These authors generalised a result of Hsu (1984b) and although, like Carroll, Gupta and Huang, they too could not derive a LFC of means,the lower bound on the PCS obtained under their method was shown to be substantially less conservative than that of Carroll, Gupta and Huang.

As discussed in Chapter 2, for the particular case of multiple comparisons with-the best ($t = 1$), Hsu (1984) used Fabian's notion of a pivotal event to construct an upper bound for

$$\theta_i - \theta_{\rho(k)}$$

of

$$\max(0, X_i - X_{R(k)} + d) \quad \forall i \neq R(k) \ .$$

Bofinger and Mengersen generalised this approach and constructed strict upper bounds for

$$\theta_i - \theta_{\rho(k-t+1)} \tag{5.9}$$

of

$$\max(0, X_i - X_{R(k-t|i)} + d) \quad \forall i = 1, .., k \tag{5.10}$$

where $X_{R(k-t|i)}$ is the tth largest of

$$X_1, X_2, .., X_{i-1}, X_{i+1}, .., X_k$$

so that

$$X_{R(k-t|i)} = \begin{cases} X_{R(k-t)} \text{ if } i = R(k-t+1), .., R(k) \\ X_{R(k-t+1)} \text{ if } i = R(1), .., R(k-t) \ . \end{cases} \tag{5.11}$$

The event that these upper bounds are correct may be written as

$$CUB = \{\theta_i - \theta_{\rho(k-t+1)} < \max(0, X_i - X_{R(k-t|i)} + d) \ \forall i = 1, .., k \ .\} \tag{5.12}$$

Bofinger and Mengersen proposed the following two lemmas to:

a) demonstrate that the event CUB is a subset of the event CS of correct selections under P1, and

b) derive a lower bound on $P_{\underline{\theta}}\{CUB\}$

and hence derived a theorem (Theorem 5.1 below) which shows that the lower bound for $P_{\underline{\theta}}\{CUB\}$ is a lower bound for $P_{\underline{\theta}}\{CS \mid P1\}$.

Lemma 5.1

Using the definitions (5.3) and (5.12) for CS and CUB respectively,

$$CUB \subseteq CS \ .$$

Proof

Since

$$\theta_i - \theta_{\rho(k-t+1)} < 0 \quad \forall i \notin \gamma_t \ ,$$

the upper bound (5.10) for $\theta_i - \theta_{\rho(k-t+1)}$ is correct $\forall i \notin \gamma_t$. Similarly,

$$\theta_i - \theta_{\rho(k-t+1)} \geq 0 \quad \forall i \in \gamma_t \ .$$

Hence the expression (5.12) may be written as

$$\begin{aligned} CUB &= \{\theta_i - \theta_{\rho(k-t+1)} < X_i - X_{R(k-t|i)} + d \ \forall i \in \gamma_t\} \qquad (5.13) \\ &\subseteq \{0 < X_i - X_{R(k-t|i)} + d \quad \forall i \in \gamma_t\} \ . \end{aligned}$$

Since

$$X_i - X_{R(k-t|i)} > 0$$

whenever $X_{R(k-t|i)} = X_{R(k-t)}$ (that is, $i = R(k-t+1), .., R(k)$), then

$$\begin{aligned} CUB &\subseteq \{0 < X_i - X_{R(k-t+1)} + d \quad \forall i \in \gamma_t\} \\ &= \{X_i > X_{R(k-t+1)} - d \quad \forall i \in \gamma_t\} \\ &= CS \ . \end{aligned}$$

Alternatively, notice that procedure P1 involves selecting all of those populations with positive upper bounds, as given by (5.10).

Lemma 5.2

$$P_{\underline{\theta}}\{CUB\} \geq (k-t+1)P_{k-1,t-1}(d) - (k-t)P_{k,t-1}(d) + P_{k-t+1,1}(d) - 1 \qquad (5.14)$$

where $P_{k,t}(d)$ is defined in (5.6).

Proof

The proof may be divided into three parts. In part a) the LFC of $\underline{\theta}$ is derived; in part b) the variable $X_{R(k-t|i)}$ is considered and in part c) the distribution functions of interest are investigated, using results from parts a) and b).

a) First the LFC of $\underline{\theta}$ is derived, being the configuration of means which minimises $P_{\underline{\theta}}\{CUB\}$. The expression for CUB (given by (5.12)) may be rearranged to obtain

$$P_{\underline{\theta}}\{CUB\} = P\{X_i - \theta_i > X_{R(k-t|i)} - \theta_{\rho(k-t+1)} - d \ \ \forall i \in \gamma_t\} \ . \qquad (5.15)$$

Notice that

i) $X_i - \theta_i$ has the same distribution for all values of $\underline{\theta}$, and

ii) $X_{R(k-t|i)}$ has the same distribution for all values of θ_i.

Consider $\theta_{\rho(k-t+1)}$ fixed. Now $X_{R(k-t|i)}$ is stochastically increasing in $\theta_j \ \forall j \neq i$, so if any θ_j is increased, $X_{R(k-t|i)}$ tends to increase and hence the probability in (5.15) is decreased.

By the ordering of the θ values,

$$\theta_i < \theta_{\rho(k-t+1)} \ \forall i \notin \gamma_t \ \text{ and } \ \theta_i \geq \theta_{\rho(k-t+1)} \ \forall i \in \gamma_t \ ,$$

so the LFC of $\underline{\theta}$ is given by

$$\theta_i \rightarrow \begin{cases} \theta_{\rho(k-t+1)} & \forall i \notin \gamma_t \\ \infty & \forall i \notin \gamma_t^* \end{cases} \qquad (5.16)$$

where

$$\gamma_t^* = \{\rho(1), \rho(2), .., \rho(k-t+1)\} \ .$$

b) Consider now $X_{R(k-t|i)}$, $i \in \gamma_t$ (since $i \notin \gamma_t$ may be ignored for the purpose of investigating (5.15)).

Under the LFC,

$$X_i \xrightarrow{a.s} \infty \ , \quad i \notin \gamma_t^* \ . \tag{5.17}$$

Hence, if $i = \rho(k-t+1)$, by considering (5.11) we can say

$$X_{R(k-t|i)} \overset{\text{a.s.}}{=} \max_{j \notin \gamma_t} X_j \ .$$

Similarly, for $i = \rho(k-t+2), .., \rho(k)$, we can say

$$X_{R(k-t|i)} \overset{\text{a.s.}}{=} \text{2nd largest}_{j \in \gamma_t^*} X_j \ .$$

Now, by writing $Y_j = X_j - \theta_j \ \forall j = 1, .., k$ and appealing to the LFC,

$$X_{R(k-t|i)} \overset{\text{a.s.}}{=} \begin{cases} \max_{j \notin \gamma_t} Y_j + \theta_{\rho(k-t+1)} & \text{if } i = \rho(k-t+1) \\ \text{2nd largest}_{j \in \gamma_t^*} Y_j + \theta_{\rho(k-t+1)} & \text{if } i \notin \gamma_t^* \ . \end{cases}$$

c) Consider now the distribution functions of interest.

i) Each Y_j has distribution function $F(y)$, $i = 1, .., k$.

ii) The distribution function for $\max_{j \notin \gamma_t} Y_j$ is given by $(F(y))^{k-t}$.

iii) The density for the 2nd largest Y_j, $j \in \gamma_t^*$, is the pdf for the $(k-t)$th order statistic from $(k-t+1)$ variables, each with cdf $F(y)$ and hence is given by

$$\begin{aligned} & \frac{(k-t+1)!}{(k-t+1-2)!(2-1)!}(F(y))^{k-t-1}(1-F(y))^{2-1}\frac{dF(y)}{dy} \\ = \ & (k-t+1)(k-t)(1-F(y))(F(y))^{k-t-1}\frac{dF(y)}{dy} \ . \end{aligned}$$

Hence at the LFC,

$$\begin{aligned} P_{\underline{\theta}}\{CUB\} &= P\{X_i - \theta_i > X_{R(k-t|i)} - \theta_{\rho(k-t+1)} - d \ \forall i \in \gamma_t\} \\ &= P\{Y_i > \text{2nd largest}_{j \in \gamma_t^*} Y_j - d \ \forall i \notin \gamma_t^* \\ &\qquad \text{and } Y_{\rho(k-t+1)} > \max_{j \notin \gamma_t} Y_j - d\} \\ &\geq P\{Y_i > \text{2nd largest}_{j \in \gamma_t^*} Y_j - d \ \forall i \notin \gamma_t^*\} \\ &\qquad + P\{Y_{\rho(k-t+1)} > \max_{j \notin \gamma_t} Y_j - d\} - 1 \ . \end{aligned}$$

Now

$$\begin{aligned}
& P\{Y_i > \text{2nd largest}_{j\in\gamma_t^*} Y_j - d \ \forall i \notin \gamma_t^*\} \\
= \ & (k-t+1)(k-t)\int (1-F(y-d))^{t-1}(1-F(y))(F(y))^{k-t-1}dF(y) \\
= \ & (k-t+1)(k-t)\int (1-F(y-d))^{t-1}(F(y))^{k-t-1}dF(y) \\
& -(k-t+1)(k-t)\int (1-F(y-d))^{t-1}(F(y))^{k-t}dF(y) \\
= \ & (k-t+1)P_{k-1,t-1}(d) - (k-t)P_{k,t-1}(d)
\end{aligned}$$

where $P_{k,t}(d)$ is defined in (5.6).

Also,

$$\begin{aligned}
& P\{Y_{\rho(k-t+1)} > \max_{j\notin\gamma_t} Y_j - d\} - 1 \\
= \ & P_{k-t+1,1}(d) \ .
\end{aligned}$$

Hence, at the LFC, the result follows.

Theorem 5.1

$$P_{\underline{\theta}}\{CS \mid P1\} \geq (k-t+1)P_{k-1,t-1}(d) - (k-t)P_{k,t-1}(d) + P_{k-t+1,1}(d) - 1 \ . \quad (5.18)$$

Proof

The proof follows from Lemmas 5.1 and 5.2.

Remark

For $t = 1$, the right hand side of expression (5.18) reduces to

$$k - (k-1) + P_{k,1}(d) - 1 = P_{k,1}(d)$$

which is Gupta's (1956) result for subset selection of the (one) best population.

Upper bound on the infimum of the PCS

Under the LFC (5.16) for the event CUB and by virtue of (5.17), it is clear that the PCS reduces to

$$P_{\underline{\theta}}\{CS \mid P1\} = P_{k-t+1,1}(d) \ . \tag{5.19}$$

This, then, is an upper bound for the infimum of $P_{\underline{\theta}}\{CS \mid P1\}$ (and hence a lower bound on the value of d such that the right hand side of (5.18) is equal to P^*.) For the normal means case, comparison of Bechhofer's (1954) tabulated values of $P_{k-t+1,1}(d)$ and Carroll, Gupta and Huang's Table III establishes the superiority of (5.19) over the latter authors' upper bound. For example, with $P^* = P_{\underline{\theta}}\{CS \mid P1\}$, for $\{P^*, k, t\} = \{0.90, 10, 2\}$, the d value satisfying (5.19) for the normal means case is 2.98, whereas the corresponding value satisfying (5.8) is 2.74. Similarly, for $\{0.95, 6, 3\}$, the new upper bound produces the value $d = 2.92$, whereas Carroll, Gupta and Huang's bound produces $d = 2.41$.

5.2.3 Normal means case

Suppose that $F(\cdot)$ is the normal distribution with mean θ and common unknown variance σ^2. Let X_i be the mean of n independent observations from π_i and let S^2 be the usual pooled estimator of σ^2, so that $\nu S^2/\sigma^2$ has a chisquared distribution on ν degrees of freedom.

Lower bound on the PCS

With goal G1 given in Section 5.2.1, we consider the selection procedure:

Procedure P2: Select a subset of populations π_i, $i \in G(d)$, where

$$G(d) = \{i : X_i > X_{R(k-t+1)} - dSn^{-1/2}\} \tag{5.20}$$

with d a positive constant.

As before, correct selection is defined to be the event (5.3), with γ_t and $G(d)$ defined by (5.1) and (5.2) respectively, and CUB becomes the event

$$CUB = \{\theta_i - \theta_{\rho(k-t+1)} < \max(0, X_i - X_{R(k-t|i)} + dSn^{-1/2}) \ \ \forall i = 1, .., k\} \ . \tag{5.21}$$

Using these definitions, Bofinger and Mengersen proved the following theorem.

Theorem 5.2

$$P_{\underline{\theta}}\{CS \mid P2\} \geq P_L, \tag{5.22}$$

where

$$P_L = (k-t+1)P^{(\nu)}_{k-1,t-1}(d) - (k-t)P^{(\nu)}_{k,t-1}(d) + P^{(\nu)}_{k-t+1,1}(d) - 1 \tag{5.23}$$

and

$$P^{(\nu)}_{k,t}(d) = t\int_0^{\infty}\int_{-\infty}^{+\infty}(\Phi(x+dS/\sigma))^{k-t}(1-\Phi(x))^{t-1}d\Phi(x)dG_\nu(S/\sigma) \tag{5.24}$$

with $\Phi(\cdot)$ the standard normal distribution function and $G_\nu(\cdot)$ the distribution function for S/σ.

Proof

The proof follows that given for Theorem 5.1, working conditionally on S and then taking the expectation over S.

Discussion of the lower bound

The sharpness of P_L as a lower bound for $P_{\underline{\theta}}\{CS \mid P2\}$ was investigated by Bofinger and Mengersen using simulation. A Monte Carlo experiment was conducted in which the NAG routines G05DHF and G05DDF were used to simulate, respectively, a sample variance S^2 and k sample means X_i distributed $N(\theta_i, 1.0)$, $i = 1,..,k$. The same set of random variables were used for each $\{P^*, k, t, \nu\}$ combination, with each combination considered a separate experiment.

Three P^* values (0.90,0.50,0.10) and two ν values (30,∞) were considered. For each of these six combinations of P^* and ν, five (k,t) combinations were considered ((k,t) = (6,2), (6,5), (10,2), (10,3), 10,8)), giving thirty $\{P^*, k, t, \nu\}$ combinations in all. Each experiment was repeated 100,000 times, giving a standard error for each simulated probability of less than 9.5E-4 for $P^* = 0.90, 0.10$ and 1.6E-3 for $P^* = 0.50$.

Table 5.1: Comparison of the PCS under configurations C1 and C2 for $\nu = \infty$.

(P^*, k, t)	$(0.90, 10, 8)$		$(0.50, 10, 3)$		$(0.10, 10, 8)$	
c	C1	C2	C1	C2	C1	C2
∞	0.958	0.958	0.738	0.738	0.679	0.679
3.0	0.961	0.971	0.749	0.760	0.684	0.704
1.0	0.980	0.997	0.813	0.851	0.728	0.826
0.5	0.984	0.998	0.807	0.838	0.725	0.818
0.1	0.982	0.993	0.772	0.795	0.699	0.792
0.0	0.981	0.981	0.757	0.757	0.689	0.689

In light of the LFC given by (5.16), two configurations of $\underline{\theta}$ were considered:

$$C1: \ \theta_i = \begin{cases} 0 & i \leq k - t + 1 \\ c & i = k - t + 2 \\ \infty & i > k - t + 2 \end{cases} \tag{5.25}$$

$$C2: \ \theta_i = \begin{cases} 0 & i \leq k - t + 1 \\ c & i \geq k - t + 2 \end{cases} \tag{5.26}$$

with $0 \leq c \leq \infty$.

As expected, for all $\{P^*, \nu, k, t\}$ combinations considered, both $P_{\underline{\theta}}\{CUB\}$ and $P_{\underline{\theta}}\{CS \mid P2\}$ were larger under configuration C2 than under configuration C1, for all values of c. This difference is illustrated in Table 5.1 for $P_{\underline{\theta}}\{CS \mid P2\}, \nu = \infty$ and three $\{P^*, k, t\}$ combinations. Hence, only configuration C1 was considered further.

For the thirty $\{P^*, k, t, \nu\}$ combinations, the simulated behaviour of $P_{\underline{\theta}}\{CS \mid P2\}$ and $P_{\underline{\theta}}\{CUB\}$ under configuration C1 was investigated over the range of c values ($0 \leq c \leq \infty$), using values of d such that $P^* = P_L$. As expected, the values of $P_{\underline{\theta}}\{CUB\}$ consistently increased in c and were less conservative than the corresponding values of $P_{\underline{\theta}}\{CS \mid P2\}$. The behaviour of $P_{\underline{\theta}}\{CS \mid P2\}$ for $\nu = \infty$ is illustrated in Figure 5.1 for $P^* = 0.90, 0.10$. The same behaviour was observed for $\nu = 30$ and $P^* = 0.50$.

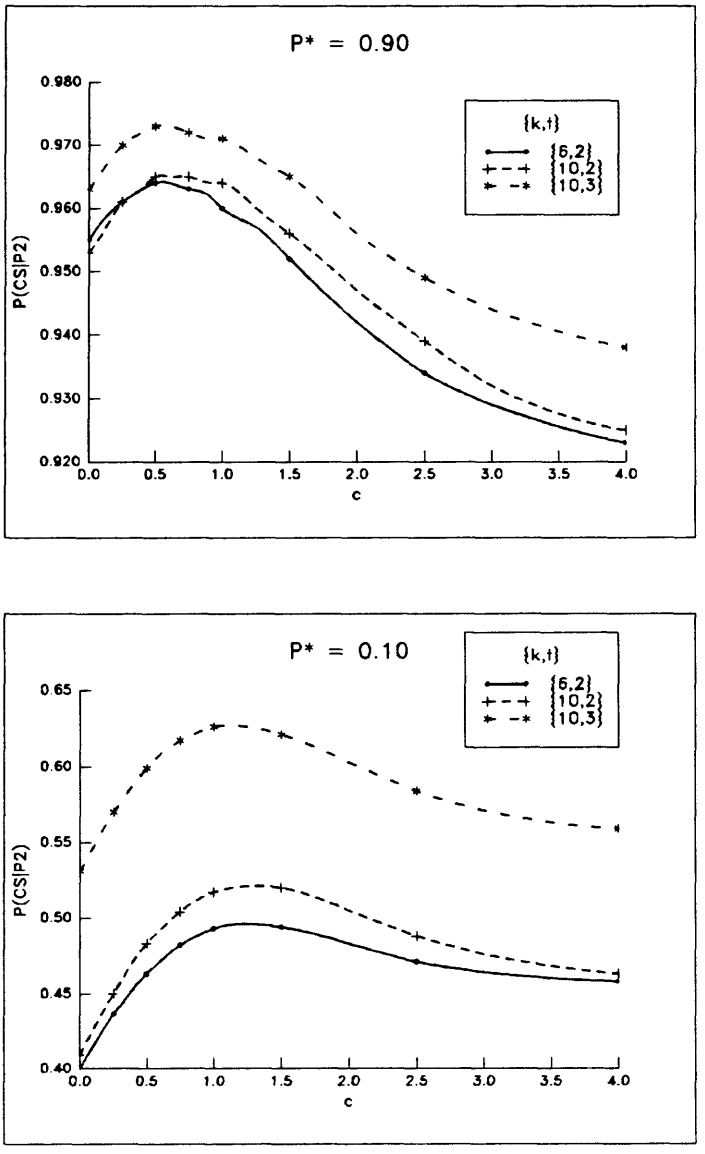

Figure 5.1: PCS at configuration $C1$ with $\nu = \infty$ and $P^* = 0.90, 0.10$

As demonstrated by Figure 5.1, the probability appeared to have a single maximum over the values of c, for all $\{P^*, k, t\}$ combinations (and for both ν values considered). The minimum probability appeared to depend on the value of P^*; for small P^* (certainly less than 0.5 according to Figure 5.1), it was at $c = 0$ while for larger P^* this minimum was at $c = \infty$. The P^* value for which equal minima were found at $c = 0$ and $c = \infty$ was dependent on the $\{\nu, k, t\}$ combination. This seems to indicate that the LFC for $P_{\underline{\theta}}\{CS \mid P2\}$ is given by (5.25), with $c = 0$ for small P^* and $c = \infty$ for larger P^* values. Hence Bofinger and Mengersen conjectured that, for sufficiently large P^* values, the LFC for $P_{\underline{\theta}}\{CUB\}$, given by (5.16), is also the LFC for $P_{\underline{\theta}}\{CS \mid P2\}$. This conjecture is discussed in more detail later in this section.

Bofinger and Mengersen also numerically compared their lower bound P_L with that of Carroll, Gupta and Huang and with the simulated values of $P_{\underline{\theta}}\{CUB\}$, for $P^* = 0.90, \nu = 30, \infty$ and various $\{k, t\}$ combinations, under configuration (5.16). These comparisons are detailed in Table 5.2. The value of $P_{\underline{\theta}}\{CS \mid P2\}$ is given by $P^{(\nu)}_{k-t+1,1}(d)$ using the appropriate value of d satisfying $P_L = 0.90$. Carroll, Gupta and Huang's probability of correct selection, $P_{CGH}\{CS\}$ say, is given by $P^{(\nu)}_{k-t+1,1}(d)$ using the value of d such that $P^{(\nu)}_{k,t}(d) = 0.90$. The latter values were computed using the package RS-MCB developed by Gupta and Hsu (1984).

Table 5.2 shows that Bofinger and Mengersen's lower bound for $P_{\underline{\theta}}\{CUB\}$ is not too far above the nominal value of 0.90. However, the lower bound for $P_{\underline{\theta}}\{CS \mid P2\}$, while an improvement on that of Carroll, Gupta and Huang, is still quite conservative, especially as k and t increase and as P^* decreases. For $P^* = 0.10$ in particular, both lower bounds on $P_{\underline{\theta}}\{CUB\}$ and $P_{\underline{\theta}}\{CS \mid P2\}$ are quite unsatisfactory. Fortunately, such probabilities are rarely used in practice (although this is no reason for not requiring further improvement in the bound on the PCS.)

Conjecture regarding the LFC Based on the simulation results discussed above, Bofinger and Mengersen conjectured that, for large P^*, the infimum of the PCS is given by $P^{(\nu)}_{k-t+1,1}(d)$. They noticed, however, that this conjecture does not hold for small P^*, as shown by the simulation results, since the LFC is not given

Table 5.2: Simulated and calculated results under configuration C1
$P\{CUB\}$ and $P\{CS \mid P2\}$ use values of d satisfying $P_L = 0.90$.
$P_{CGH}\{CS\}$ uses values of d satisfying $P_{k,t}^{(\nu)}(d) = 0.90$.

$\nu = \infty$					$\nu = 30$			
t	k	$P\{CUB\}$	$P\{CS\}$	$P_{CGH}\{CS\}$	t	k	$P\{CUB\}$	$P\{CS\}$	$P_{CGH}\{CS\}$
2	4	0.900	0.922	0.944	2	4	0.903	0.923	0.943
	8	0.900	0.924	0.946		8	0.903	0.925	0.944
	10	0.902	0.925	0.946		10	0.903	0.926	0.944
3	6	0.902	0.935	0.961	3	6	0.904	0.936	0.960
	10	0.903	0.937	0.962		10	0.904	0.939	0.961
5	6	0.900	0.947	0.972	5	6	0.900	0.948	0.971
	10	0.903	0.950	0.975		10	0.906	0.951	0.974
8	10	0.903	0.958	0.982	8	10	0.904	0.959	0.981

by (5.16) in these cases. For large P^*, however, it was pointed out that if the conjecture were true, the percentage points d would be reduced. For example, for $\{P^*, \nu, k, t\} = \{0.75, \infty, 10, 5\}$, the value of $d = 2.63$ satisfying $P_L = 0.75$ would be replaced by 1.97 (satisfying $P_{k-t+1,1}^{(\nu)}(d) = 0.75$) and, for $\{0.95, 30, 6, 3\}$, d would decrease from 3.30 to 3.03. (Notice that the values of d appropriate to Carroll, Gupta and Huang's procedure for these combinations would be 2.94 and 3.66 respectively.)

Bofinger (1988) has shown that this conjecture does hold for $t = k - 1$. Hence, for $(k, k-1)$ combinations, the infimum of the PCS is given by $P_{2,1}^{(\nu)}(d)$.

Tables for the lower bound on the PCS - normal means case

Computation of the expression for $P_{k,t}^{(\nu)}(d)$, given by (5.24), was discussed in Chapter 4. As indicated in that chapter, a satisfactory approximation to $P_{k,t}^{(\nu)}(d)$ may be obtained by writing (5.24) as

$$P_{k,t}^{(\nu)}(d) = 1 - \int_0^1 G_\nu(S/\sigma) dP_{k,t}(dS/\sigma)$$

and using Bechhofer's (1954) tables to approximate $P_{k,t}(dS/\sigma)$, the inner integral of (5.24). This approach is much faster than alternatives based on Gaussian quadrature and gives percentage points accurate to one in the second decimal place. The

method was again employed to compute values of d such that P_L is within a small amount $\delta = 2.0E-5$ of $P^* = 0.75, 0.90, 0.95$. The results are presented in Table B.3 for selected values of ν, k and t. (This table is located in Appendix B.) Using Bofinger's (1988) results, the values of d in Table B.3 for $t = k-1$ are those satisfying $P^* = P_{2,1}^{(\nu)}(d)$.

As an overall check on the consistency of the values presented in Table B.3, a response surface was fitted over k and ν to each (P^*, t) combination, as described in Chapter 4. The outliers were verified by computing the percentage points of $P_{k,t}^{(\nu)}(d)$ using the routine Pktnud described in Chapter 4. None of the tabulated values, when compared with the new percentage points, differed by more than 0.01 and always the tabulated values were the more conservative.

Interpolation For values of ν not included in Table B.3, and for P^* values between 0.75 and 0.95, the interpolation methods suggested for Table B.1 (described in Chapter 4) apply. The behaviour of these methods is similar to that described in Chapter 4 for Table B.1.

5.3 Selection of all t good populations - scale Parameter Case

Consider k independent random variables $Y_i, i = 1, .., k$, from populations π_i with continuous distribution functions $F(y/\psi_i)$. We order the ψ_i and Y_i as follows:

$$\psi_{\rho(1)} \leq \psi_{\rho(2)} \leq \ldots \leq \psi_{\rho(t)} < \psi_{\rho(t+1)} \leq \ldots \leq \psi_{\rho(k)} \ ;$$

$$Y_{R(1)} < Y_{R(2)} < \ldots < Y_{R(k)} \ .$$

The selection goal is now to select a subset, based on the Y_i values, of "good" populations, with a good population now identified as possessing one of the t smallest ψ_i values.

Goal G3: With PCS at least P^*, select all populations $\pi_i, i \in \gamma_{t'}$, where

$$\gamma_{t'} = \{\rho(1), \rho(2), .., \rho(t)\} \ . \tag{5.27}$$

This problem of selecting populations based on scale parameters has been considered under the subset selection approach by Gupta and Sobel (1962) for the normal variances case, and by Carroll, Gupta and Huang (1975) and Mengersen (1987) for more general distributions. In the latter paper, Mengersen extended in an obvious way the results obtained by Bofinger and Mengersen for the location parameter case, described in Section 5.2. Mengersen's results are detailed in this section.

In all three papers, the selection procedure is:

Procedure P3: To select all populations π_i , $i \in G(c)$, $\forall i \in G(c)$, where

$$G(c) = \{i : Y_i < c^{-1} Y_{R(t)}\} ,$$

with $0 < c < 1$ constant. (This procedure is considered in Section 7.2 of Gupta and Sobel's paper and corresponds to procedures R_4 of Carroll, Gupta and Huang and Q1 of Mengersen.)

A correct selection then becomes the event

$$CS = \{\gamma_{t'} \subseteq G(c)\}$$

which, again, is the event that all t good populations are selected.

5.3.1 A previous approach

For the scale parameter case, Carroll, Gupta and Huang adopted the same approach as that taken for the location parameter problem. Since they were unable to find a LFC for $\underline{\psi}$ (such that $P_{\underline{\psi}}\{CS \mid P3\}$ reaches an infimum), Carroll, Gupta and Huang proposed the following lower bound:

$$\inf P_{\underline{\psi}}\{CS \mid P3\} \geq Q_{k,t}^{(\nu)}(c)$$

where

$$Q_{k,t}^{(\nu)}(c) = t \int_0^\infty (1 - F(yc))^{k-t} ((F(y))^{t-1} dF(y) . \tag{5.28}$$

Gupta and Sobel (1962) also proposed this bound for the normal variances case.

5.3.2 A new approach

Bofinger and Mengersen's procedure for the selection of the t best populations in the location parameter case was extended by Mengersen to the scale parameter problem. Strict lower bounds on

$$\psi_i/\psi_{\rho(t)}$$

$\forall i = 1, .., k$ of

$$\min(1, cY_i/Y_{R(t|i)}) \tag{5.29}$$

were constructed, where $Y_{R(t|i)}$ is the tth smallest of

$$Y_1, Y_2, .., Y_{i-1}, Y_{i+1}, .., Y_k \ .$$

Then the selection procedure P3 is equivalent to selecting all populations with lower bounds (5.29) less than 1.

Theorem 5.3

$$P_{\underline{\psi}}\{CS \mid P3\} \geq (k-t+1)Q^{(\nu)}_{k-1,t-1}(c) - (k-t)Q^{(\nu)}_{k,t-1}(c) + Q^{(\nu)}_{k-t+1,1}(c) - 1 \tag{5.30}$$

where $Q^{(\nu)}_{k,t}(c)$ is defined in (5.28).

Proof

Let $X = -\log Y$ and $d = -\log c$. Then the distribution functions become location-type functions $F(x - \theta_i)$ with

$$\theta_i = -\log \psi_i \ .$$

Also, procedure P3 is equivalent to selecting π_i if

$$-\log Y_i > -\log Y_{R(t)} + \log c \ ,$$

that is, if

$$X_i > t\text{th largest } X - d \ ,$$

which is procedure P1 described in Section 5.2.

Hence we may apply the proofs of Lemmas 5.1 and 5.2 and Theorem 5.1 to obtain the required result.

5.3.3 Normal variances case

Let Y_i be the sample variance based on n observations from each of the normally distributed populations π_i, $i = 1, .., k$. Then the distribution function $F(y/\psi_i)$ is that of a chisquared variable on $\nu = n - 1$ degrees of freedom and $\frac{\nu Y_i}{\sigma^2}$ is distributed as chisquared on ν degrees of freedom.

For this particular case, the values of c such that the right hand side of expression (5.30) is within a small value $\delta = 2.0E - 5$ of $P^* = 0.75, 0.90, 0.95$ were presented by Mengersen (1987) and are displayed in Table B.4 (see Appendix B) for various $\{\nu, k, t\}$ combinations. The method of computing expression (5.28) is described in Section 4.3. It is expected that the tabulated values are correct to within one in the second decimal place.

By comparing the values of c in Table B.4 with those of Table B.2 (for $Q_{k,t}^{(\nu)}(c)$), for given $\{P^*, \nu, k, t\}$ it can be seen that the new procedure is an improvement over that of Carroll, Gupta and Huang (since larger c values are obtained in the former case). For example, the value of c in Table B.4 corresponding to $\{P^*, \nu, k, t\} = \{0.75, 8, 5, 2\}$ is 0.346, whereas the corresponding value under Carroll, Gupta and Huang's procedure (from Table B.2) is 0.319. Similarly, Mengersen's value of c for $\{0.95, 60, 10, 3\}$ is 0.520 whereas Carroll, Gupta and Huang's value is 0.488.

The overall consistency of the tabulated values was checked in the same way as for Table B.3, by fitting a response surface and examining residuals. Those values corresponding to large residuals were recomputed and, where possible, were compared with Carroll, Gupta and Huang's or Gupta and Sobel's tabulated values.

Interpolation The methods of interpolation suggested for Table B.2 apply to Table B.4, for ν values not included in the table and for P^* values between 0.75 and 0.95.

Results for large degrees of freedom For large degrees of freedom, Mengersen considered the use of the normal approximation to the chisquared distribution, to derive the required values of c using the corresponding values of d in Table B.3 (for the location parameter case). Pairs of values of c (obtained from Table B.4)

Table 5.3: Comparison of c and $\hat{c}$
c is value satisfying $P_L = P^*$
$\hat{c}$ is value satisfying normal approximation.

P^*	ν		(k,t)								
			(4,2)	(7,2)	(10,2)	(4,3)	(7,3)	(10,3)	(7,5)	(10,5)	(10,7)
0.75	10	c	0.443	0.336	0.293	0.518	0.338	0.287	0.389	0.296	0.330
		$\hat{c}$	0.412	0.311	0.272	0.475	0.302	0.256	0.340	0.252	0.277
	30	c	0.632	0.545	0.509	0.686	0.544	0.500	0.583	0.504	0.531
		$\hat{c}$	0.623	0.540	0.505	0.671	0.534	0.492	0.566	0.490	0.512
	60	c	0.724	0.655	0.624	0.766	0.653	0.616	0.683	0.618	0.640
		$\hat{c}$	0.719	0.654	0.624	0.759	0.648	0.613	0.674	0.611	0.630
	120	c	0.796	0.743	0.719	0.828	0.740	0.711	0.763	0.712	0.729
		$\hat{c}$	0.794	0.743	0.721	0.824	0.737	0.710	0.760	0.709	0.724
0.95	10	c	0.263	0.205	0.181	0.312	0.212	0.183	0.247	0.193	0.216
		$\hat{c}$	0.215	0.161	0.139	0.255	0.160	0.134	0.184	0.134	0.361
	30	c	0.475	0.418	0.393	0.518	0.424	0.393	0.456	0.401	0.423
		$\hat{c}$	0.460	0.402	0.377	0.500	0.404	0.374	0.430	0.375	0.394
	60	c	0.594	0.544	0.521	0.629	0.548	0.520	0.575	0.527	0.546
		$\hat{c}$	0.587	0.538	0.515	0.620	0.539	0.512	0.563	0.514	0.531
	120	c	0.728	0.652	0.633	0.720	0.655	0.632	0.676	0.637	0.652
		$\hat{c}$	0.690	0.652	0.631	0.717	0.651	0.629	0.671	0.631	0.645

and $\hat{c}$ (obtained using the normal approximation) for various $\{P^*, \nu, k, t\}$ combinations are given in Table 5.3 for illustration. As expected, the deficiencies of this approximation that were noticed in Chapter 4 were more apparent for the results of the present chapter. It appears that the usual normal approximation may not give satisfactory results for this selection goal for $t > 1$. As noticed in Chapter 4, a better method of approximating the c values for large degrees of freedom, for general values of t, remains to be investigated.

A conjectured lower bound

For the normal variances case, Gupta and Sobel (1962) conjectured that the LFC of $\underline{\psi}$ is given by

$$\psi_{\rho(i)} = \begin{cases} 0 & i = 1, .., s-1 \\ \psi_{\rho(k)} & i = s, .., k-1 \end{cases}$$

where $s(1 \leq s \leq t)$ is a non-decreasing function of P^*. Hence, for large P^*, s is set equal to t and the required value of c is that value satisfying $Q^{(\nu)}_{k-t+1,1}(c) = P^*$. This is analagous to the conjecture proposed by Bofinger and Mengersen (1986) for the location parameter case, which is described in Section 5.2. Gupta and Sobel proved their conjecture for the special case $\{\nu, k, t\} = \{2, 3, 2\}$. For this case, the exact value of c, for P^* between $1/3$ and $2/3$, is shown to be

$$c = (3(1 - P^*) + \sqrt{\{(1 - P^*)(17 - P^*)\}})/(4(1 + P^*))$$

and, for P^* between $2/3$ and 1, the value is given by

$$c = (1 - P^*)/P^* \ .$$

If this conjecture were true for all values of ν, k and t, then the critical values of c appropriate to Procedure P3 could be considerably reduced. For example, for $\{P^*, \nu, k, t\} = \{0.75, 8, 5, 2\}$ the value of c would be 0.398, compared with Mengersen's value of 0.346 (from Table B.4) and Carroll, Gupta and Huang's value 0.319 (from Table B.2). For $\{0.95, 60, 10, 3\}$, c would be 0.537, compared with 0.520 from Table B.4 and 0.488 from Table B.2.

5.4 Selecting only good populations

5.4.1 Location parameter case

As well as selecting a subset containing *all* good populations with a prespecified PCS, Bofinger and Mengersen considered the problem of selecting a subset containing *only* good populations. The goal in this case becomes:

Goal G4: Select only those π_i, $i \in \gamma_t$, where γ_t is defined by (5.1). The corresponding procedure proposed by Bofinger and Mengersen is:

Procedure P4: Select all π_i, $i \in G^*(d)$, where

$$G^*(d) = \{i : X_i > X_{R(k-t)} + d\} \tag{5.31}$$

with d a positive constant satisfying the probability requirement

$$P_{\underline{\theta}}\{CS \mid P4\} = P_{\underline{\theta}}\{G^*(d) \subseteq \gamma_t\} \ . \tag{5.32}$$

Taking an approach similar to that described in Section 5.2, Bofinger and Mengersen constructed strict *lower* bounds for

$$\theta_i - \theta_{\rho(k-t)}$$

of

$$\min(0, X_i - X_{R(k-t|i)} - d) \quad \forall i = 1, .., k \tag{5.33}$$

with $X_{R(k-t|i)}$ defined in (5.11).

The event that these lower bounds are correct is given by

$$CLB = \{\theta_i - \theta_{\rho(k-t)} > \max(0, X_i - X_{R(k-t|i)} - d) \ \forall i = 1, .., k \ .\} \tag{5.34}$$

As in Section 5.2, a lower bound is obtained on $P_{\underline{\theta}}\{CS \mid P4\}$ by:

i) establishing that the event CLB is a subset of the event CS of correct selections, and

ii) finding a lower bound for $P_{\underline{\theta}}\{CLB\}$.

i) With the definitions of CS and CLB given by (5.3) and (5.34) respectively, an obvious modification of the proof of Lemma 5.1 may be made to show that

$$CLB \subseteq CS \ .$$

This may alternatively be demonstrated by noticing that Procedure P4 involves selecting all populations with non-negative lower bounds, as given by (5.33).

ii) Selection of *only* "good" populations (those $\pi_i, i \in \gamma_t$) is equivalent to rejection of *all* $(k - t)$ "bad" populations (those π_i, $i \notin \gamma_t$). Hence, a lower bound for $P_{\underline{\theta}}\{CLB\}$ may be obtained by **replacing** t with $(k - t)$ in the right hand side of expression (5.18). Bofinger and Mengersen thus showed that

$$P_{\underline{\theta}}\{CLB\} \geq (t+1)P_{k-1,t}(d) - tP_{k,t+1}(d) + P_{t+1,1}(d) - 1 \ . \tag{5.35}$$

For the case of normal means with common unknown variance, these values may be obtained for a particular $\{P^*, \nu, k, t\}$ combination from Table B.3, using the value corresponding to $\{P^*, \nu, k, k - t\}$.

5.4.2 Scale parameter case

Selection of a subset containing only good populations, with a good population π_i in this case possessing one of the t smallest scale parameters ψ_i, was considered by Mengersen (1987) using an extension of Bofinger and Mengersen's (1986) results. The selection goal considered in this case is:

Goal G5: Select all π_i, $i \in \gamma_{t'}$, where $\gamma_{t'}$ is given by (5.27)

and the corresponding procedure is, $\forall i = 1, .., k$:

Procedure P5: To select those π_i, $i \in G^*(c)$, where

$$G^*(c) = \{ i : Y_i < c^{-1} Y_{R(t+1)} \}$$

with probability of correct selection satisfying

$$P_{\underline{\psi}}\{CS \mid P5\} = P\{G^*(c) \in \gamma_{t'}\} .$$

Strict upper bounds may be constructed for

$$\psi_i / \psi_{\rho(t+1)}$$

$\forall i = 1, .., k$ of

$$\max(1, cY_i / Y_{R(t+1|i)})$$

with $Y_{R(t+1|i)}$ defined as the $(t+1)$th smallest of

$$Y_1, Y_2, .., Y_{i-1}, Y_{i+1}, .., Y_k .$$

The following points may be made:

i) The procedure P5 corresponds to selecting those populations with upper bounds equal to 1. Hence the event that these upper bounds are correct is a subset of the event of correct selection.

ii) A lower bound on $P_{\underline{\psi}}\{CS \mid P4\}$ is given by

$$P_{\underline{\psi}}\{CS \mid P4\} \geq (t+1) Q^{(\nu)}_{k-1,t}(c) - t Q^{(\nu)}_{k,t+1}(c) + Q^{(\nu)}_{t+1,1}(c) - 1 .$$

As for the location parameter case, this lower bound is obtained by replacing t with $(k-t)$ in expression (5.30) (which is a lower bound on correct selection of *all* t good populations and hence *only* $(k-t)$ bad populations).

5.5 Example

Consider the example, given in Chapter 3, of eight candidates (in this case, lecturers) evaluated by 21 judges (students) for particular awards, with two of these candidates eliminated at an earlier stage. The final scores for the six remaining candidates are given by:

Carole	Debbie	Wayne	Wendy	Russell	Janice
134.7	125.2	118.7	111.5	89.5	70.2

with s^2 equal to 107.04 on 140 degrees of freedom.

As proposed in Chapter 3, these students may wish to select two "best" lecturers, that is, the lecturers with the largest true scores $\theta_{\rho(6)}$ and $\theta_{\rho(5)}$.

Suppose that they wish to achieve this with 90% confidence of correct selection. From Table B.3, with $\{P^*, \nu, k, t\} = \{0.90, 140, 6, 2\}$, $d = 2.81$. Hence $dsn^{-1/2} = 6.35$ and the students may select Carole and Debbie, with 90% confidence that these are indeed the best lecturers with respect to the specified attribute . At the 95% confidence level, however, the value of d from Table B.3 is 3.22 and the set of lecturers {Carole, Debbie, Wayne} must be chosen in order to be 95% confident that the two best lecturers are included.

If the students decided instead to select *only* good lecturers, then at the 90% level, Carole and Debbie would again be chosen. At the 95% level, however, only Carole could be chosen.

It is interesting to compare these selection decisions based on Procedure P2 with those resulting from Carroll, Gupta and Huang's procedure R_3 and those possible if the conjecture in the previous section were true (procedure P_{conj}, say). Selections under each of these procedures are illustrated in Table 5.4.

It can be seen that Bofinger and Mengersen's procedure in this example outperforms Carroll, Gupta and Huang's procedure, but could be much improved if the former authors' conjecture were proven to be true for all k.

Table 5.4: Comparison of selections under Procedures $R_3, P2$ and P_{conj}.

P^*	Procedure	Select *all* good lecturers	Select *only* good lecturers
0.90	R_3	Carole,Debbie,Wayne	Carole
	P2	Carole,Debbie	Carole,Debbie
	P_{conj}	Carole,Debbie	Carole,Debbie
0.95	R_3	Carole,Debbie,Wayne	Carole
	P2	Carole,Debbie,Wayne	Carole
	P_{conj}	Carole,Debbie	Carole,Debbie

Chapter 6

SIMULTANEOUS COMPARISONS WITH A CONTROL AND WITH THE BEST

6.1 Introduction

In the previous two chapters, emphasis has been on selecting "best" populations, with respect to a particular population parameter. In many experiments, however, comparisons are not merely required with the best population, but also with some known or unknown control. The experimenter in this situation may wish to use such comparisons in order to select populations which are demonstrably better than the control *and* not demonstrably worse than the (unknown) "best" population. If no such population exists, the experimenter may wish to select only the control. In any case, comparisons are required which provide useful information about the relative performance of the experimental populations compared with the control and with the best.

A possible approach that the experimenter may take is to include the control population with the experimental populations and apply Hsu's (1984) confidence

bound procedure for comparison of all populations with the (unknown) best. These intervals may then be used to select a subset containing the best population. If the control is included in the selected subset, however, the only conclusion that the experimenter may make is that no population is demonstrably better than the control. If the control is not in the selected subset, it may be concluded that at least one of the selected populations is demonstrably better than the control, but the experimenter will be unable to identify the particular populations.

Alternatively, Dunnett's (1955,1964) one- and two-sided confidence intervals may be employed to compare all the experimental populations with the control and hence to select a subset of populations which are demonstrably better than the control. This approach, however, reveals no information about the best population. Brostrom's (1981) step-down procedures are also useful for comparisons with a control, but again yield no comparison with the best.

Simultaneous comparison of populations with a control and with the best has received attention recently from such authors as Turnbull (1976), Bechhofer and Turnbull (1978) and Bristol and Desu (1985). In all of these papers, the indifference zone approach to selection was employed. For the location parameter case, in which a population π_i is declared to be "better" than π_j if the corresponding location parameter θ_i is larger than θ_j, all three methods require the the experimenter to specify five constants $\{\delta_1, \delta_2, \delta_3, P_0, P_1\}$ such that:

(i) the probability of selecting no population as demonstrably better than the control is at least P_0 whenever $\theta_{\rho(k)} - \theta_0 < \delta_1$, where $\theta_{\rho(k)}$ is the parameter corresponding to the (unknown) best population and θ_0 is the parameter corresponding to the control, and

(ii) the probability of correctly selecting the best population is at least P_1 whenever $\theta_{\rho(k)} - \theta_0 \geq \delta_2$ *and* $\theta_{\rho(k)} - \theta_{\rho(k-1)} \geq \delta_3$.

All three methods further assume that the parameter corresponding to the control population is fixed and known. Bechhofer and Turnbull (1978) considered only the normal means case, whereas Bristol and Desu's (1986) method is designed for the case of comparing guarantee times of two-parameter exponential distributions.

If the scale parameter is common and known, both sets of authors proposed a one-stage procedure, but if the scale parameter is common and unknown, two-stage procedures were recommended. Finally, if no populations are found to be better than the control, the only conclusion that may be made under these rules is that "no populations are selected".

There are a number of disadvantages with these methods.

- Often the experimenter cannot, or is reluctant to, specify the five parameters required for implementation of the procedures.
- In many cases the control mean may be unknown.
- The scale parameters of the populations may be unknown but a two- stage experiment may be impractical.
- More importantly, however, the experimenter may require more detailed information about the comparative performance of the populations. Results expressed as numerical comparisons between experimental populations and the control, and between experimental populations and the best, may be required rather than merely the final selection statement. If a population is of interest for other reasons (such as cost efficiency) and can be shown to be "not too far from best" or "not much worse than the control", this may be useful to the experimenter.

An alternative approach to this problem has been proposed by Bofinger and Mengersen (1988). Using a confidence bound approach, bounds were obtained on comparisons of all experimental populations with the control. For those populations found to be demonstrably better than the control, intervals were constructed for comparisons with the (unknown) best population. It is these results that are discussed in detail in this chapter. The bounds proposed by Bofinger and Mengersen are detailed in Section 6.2 and in Section 6.3 various selection decisions based on the comparisons are considered. Three cases are identified:

(i) comparisons with the control are required to be as sensitive as possible.

(ii) comparisons with the best are required to be as sensitive as possible, and

(iii) comparisons with the control and the best are required to be equally sensitive.

Bofinger and Mengersen considered in detail the special case of normal means with common unknown variance and produced appropriate tables for each of the three cases above. These results are presented in Section 6.4, with an example in Section 6.6. In Section 6.5 two related problems are considered: (i) simultaneous comparisons with the control and with the t best and (ii) selection of the best provided that its mean does not lie in a prespecified "undesirable" region.

6.2 Confidence Bounds

6.2.1 Notation

Let $X_i, i = 1, .., p$, be random variables from populations π_i with continuous distribution functions (cdf) $F_1(x - \theta_i)$. Let X_0 be a random variable from π_0 , the control population with which the other p populations are to compared, with cdf $F_0(x - \theta_0)$, θ_0 unknown. If θ_0 is known (so that a known standard is employed instead of a control), then $F_0(x - \theta_0)$ is replaced by a degenerate distribution concentrated at the known θ_0.

Let $\underline{\theta} = (\theta_0, \theta_1, ..., \theta_p)$ and let Ω be the admissible space for $\underline{\theta}$.

The θ and X values are ordered as follows:

$$\theta_{\rho(1)} \leq \theta_{\rho(2)} \leq \ldots \leq \theta_{\rho(p-1)} < \theta_{\rho(p)} \tag{6.1}$$

$$X_{R(1)} < X_{R(2)} < \ldots < X_{R(p)} \tag{6.2}$$

where $\rho(\cdot)$ and $R(\cdot)$ are parametric and random permutation functions respectively. The population $\pi_{\rho(p)}$ is then defined as the true best population (excluding the control).

The set of populations which are truly better than the control is denoted by $\{\Pi_i\}, i \varepsilon \gamma$, where

$$\gamma = \{i : \theta_i - \theta_0 > 0\} \tag{6.3}$$

with $\#\gamma = g$ and with the complementary set given by

$$\overline{\gamma} = \{1, 2, .., p\} - \gamma \ .$$

6.2.2 Construction of bounds

Bofinger and Mengersen (1988) identified those populations $\pi_i, i = 1, .., p$ which are observed to be demonstrably better than the control by the set:

$$G = \{i : X_i - X_0 > c\}, \quad c > 0 \tag{6.4}$$

with

$$\overline{G} = \{1, 2, ..., p\} - G$$

the complementary set. These populations may also be identified by constructing lower bounds on

$$\theta_i - \theta_0$$

$\forall i = 1, .., p$ of

$$L_i = \min(0, X_i - X_0 - c) \ . \tag{6.5}$$

Those $\pi_i, i \in G$, will have zero (rather than negative) lower bounds and those $\pi_i, i \in \overline{G}$, will have negative lower bounds. In Bofinger and Mengersen's procedure, the latter populations are assumed to be of little interest to the experimenter, so are not compared with the best population.

For $\pi_i, i \in G$, intervals for

$$\theta_i - \max_{j \neq i, 0} \theta_j$$

of

$$I_i = [\min(0, X_i - \max_{j \neq i, 0} X_j - b), \max(0, X_i - \max_{j \in G, \neq i} X_j + b)] \tag{6.6}$$

were constructed by Bofinger and Mengersen, with b another positive constant.

No upper bound for the interval I_i is defined if $\#G = 1$ (since $\max_{j \in G, \neq i} X_j$ is undefined). Bofinger and Mengersen noted that if such an upper bound is attempted, perhaps by using $\max_{j \neq 0, i} X_j$ instead of $\max_{j \in G, \neq i} X_j$, a smaller event than E_b (given below by (6.9)) must be considered, resulting in reduced confidence if the same c and b values are used.

The joint confidence for the bounds given by (6.5) and (6.6) is given in Bofinger and Mengersen's Theorem 2.1, which is detailed below in Theorem 6.1.

Theorem 6.1

With G, L_i and I_i defined in (6.4), (6.5) and (6.6) respectively, the joint confidence for the bounds L_i and I_i is given by

$$\begin{aligned} & \inf_{\underline{\theta} \in \Omega} P\{\theta_i - \theta_0 > L_i \ \forall i = 1, .., p \ \text{ and } \ \theta_i - \max_{j \neq i, 0} \theta_j \in I_i \ \forall i \in G\} \\ = \ & \min_{0 \leq r \leq p} V_{p-r+1}(c) U_r(b) \end{aligned} \tag{6.7}$$

with

$$\left.\begin{aligned} U_0(b) &= 1 \\ U_r(b) &= \int (F_1(x+b))^{r-1} dF_1(x) \\ V_{p-r+1}(c) &= \int (F_1(x+c))^{p-r} dF_0(x) \ . \end{aligned}\right\} \tag{6.8}$$

Proof

Let independent events E_c and E_b be expressed as:

$$\left.\begin{aligned} E_c &= \{X_i - \theta_i < X_0 - \theta_0 + c \ \forall i \in \overline{\gamma}\} \\ E_b &= \{X_i - \theta_i < X_{\rho(p)} - \theta_{\rho(p)} + b \ \forall i \in \gamma, \neq \rho(p)\} \end{aligned}\right\} \tag{6.9}$$

so that

$$\left.\begin{aligned} P\{E_c\} &= V_{p-g+1}(c) \ , \\ P\{E_b\} &= U_g(b) \ . \end{aligned}\right\} \tag{6.10}$$

Notice that

$$\theta_i - \theta_0 > 0 \ \ \forall i \in \gamma \ .$$

Hence, if the event E_c occurs, then for all p populations, L_i is a correct lower bound for $\theta_i - \theta_0$; that is,

$$E_c \Rightarrow \{\theta_i - \theta_0 > L_i \ \ \forall i = 1, .., p\} \ .$$

We need, then, to show that

$$E_c \cap E_b \Rightarrow \{\theta_i - \max_{j \neq i, 0} \theta_j \in I_i \ \ \forall i \in G\} \ . \tag{6.11}$$

Notice the following:

(i) $\theta_{\rho(p)} - \max_{j \neq \rho(p),0} \theta_j > 0$

and hence the lower bound given by I_i is correct for $i = \rho(p)$.

(ii) $\theta_i - \theta_{\rho(p)} < 0 \quad \forall i \neq \rho(p), 0$

and hence the upper bounds given by I_i are correct for $i \neq \rho(p)$.

(iii) $E_c \Rightarrow \{G \subseteq \gamma\}$

and

$$\max_{j \neq i,0} X_j \geq X_{\rho(p)} \quad \forall i \neq \rho(p) \ ,$$

hence

$$\begin{aligned} E_c \cap E_b &\Rightarrow \{\theta_i - \theta_{\rho(p)} > X_i - X_{\rho(p)} - b \ \forall i \in G, \neq \rho(p)\} \\ &\Rightarrow \{\theta_i - \max_{j \neq i,0} \theta_j > X_i - \max_{j \neq i.0} X_j - b \ \forall i \in G, \neq \rho(p)\} \ . \end{aligned} \tag{6.12}$$

(iv) If $\#G > 1$, then with $X_M = \max_{j \in G, \neq \rho(p)} X_j$,

$$\begin{aligned} E_c \cap E_b &\Rightarrow \{\theta_{\rho(p)} - \theta_j < X_{\rho(p)} - X_j + \ b \ \forall j \in G, \neq \rho(p)\} \\ &\Rightarrow \{\theta_{\rho(p)} - \theta_M < X_{\rho(p)} - X_M + b\} \\ &\Rightarrow \{\theta_{\rho(p)} - \theta_{\rho(p-1)} < X_{\rho(p)} - X_M + b\} \ . \end{aligned} \tag{6.13}$$

From (6.12) and (6.13) we find that

$$\begin{aligned} P\{ \ & \theta_i - \theta_0 > L_i \ \forall i = 1, .., p \ \text{ and} \\ & \theta_i - \max_{j \neq i,0} \theta_j \in I_i \ \forall i \in G\} \\ \geq \ & P\{E_c \cap E_b\} \\ = \ & V_{p-g+1}(c) U_g(b) \\ \geq \ & \min_{0 \leq r \leq p} V_{p-r+1}(c) U_r(b) \ . \end{aligned}$$

Now consider a configuration of $\underline{\theta}$ given by

$$\begin{aligned} \theta_{\rho(p-r)} &\leq \theta_0, \quad r \geq 2 \\ \theta_{\rho(p-r+1)} &= \theta_{\rho(p-1)} \to \theta_{\rho(p)} \to \infty \ . \end{aligned}$$

Under this configuration,

$$P\{\theta_i - \theta_0 > L_i \ \ \forall i = 1,..,p\} \rightarrow V_{p-r+1}(c) \ .$$

Then conditionally on $G \stackrel{\text{a.s.}}{=} \gamma$,

$$\begin{aligned} & P\{\theta_i - \max_{j \neq i,0} \theta_j \in I_i \ \forall i \in G \mid \theta_i - \theta_0 > L_i \ \forall i = 1,..,p\} \qquad (6.14) \\ \rightarrow \ & P\{0 > X_i - \max_{j \in \gamma, \neq i} X_j - b \ \forall i \in \gamma, \neq \rho(p) \ \text{ and} \\ & 0 < X_{\rho(p)} - \max_{j \in \gamma, \neq \rho(p)} X_j + b\} \end{aligned}$$

since $r \geq 2$.

Taking the second part of this expression,

$$\begin{aligned} & \{X_{\rho(p)} > \max_{j \in \gamma, \neq \rho(p)} X_j - b\} \\ = \ & \{X_i - b < X_{\rho(p)} \ \forall i \in \gamma, \neq \rho(p)\} \\ \Rightarrow \ & \{X_i - b < \max_{j \in \gamma, \neq i} X_j \ \forall i \in \gamma, \neq \rho(p)\} \end{aligned}$$

and hence the conditional probability above approaches

$$\begin{aligned} & P\{X_{\rho(p)} > \max_{j \in \gamma, \neq \rho(p)} X_j - b\} \\ = \ & U_r(b) \ . \end{aligned}$$

For $r \geq 2$, then, the result is proven. If the minimum in (6.7) in fact occurs at $r = 0$, the infimum is attained at a configuration in which

$$\theta_{\rho(p)} < \theta_0 \ .$$

The minimum cannot occur at $r = 1$ since

$$V_p(c)U_1(b) = V_p(c) > V_{p+1}(c)U_0(b)$$

which is the value at $r = 0$.

Remark

Bofinger and Mengersen noted that, for $i \in G$ and $\#G > 1$, the intervals I_i are given by Hsu (1984) for $i = 1, .., p$. Hsu's intervals are preferred if *no* comparisons with the control are required, since the confidence is generally larger with the same values of b and c or, alternatively, smaller (b, c) combinations may be used with the same confidence. Also, more comparisons are made in this case. It is shown later, however, that if no comparisons with the control are required, Bofinger and Mengersen's procedure uses the minimum possible value of b, which is the same as Hsu's value and which results in the same confidence as that of Hsu.

Bofinger and Mengersen gave the following lemma and theorem for use in calculating the probability (6.7). (The proofs for these may be found in the paper by Bofinger and Mengersen (1988).)

Lemma 6.1

The expression

$$Q_r(c, b) = V_{p-r+1}(c)U_r(b) \tag{6.15}$$

is strictly convex in r, $\forall r \in [2, p-1]$. Also

$$Q_0(c, b) < Q_1(c, b) \ .$$

Theorem 6.2

a) If $\exists r_M \in [2, p-1]$ such that

$$Q_{r_M}(c, b) \leq \min(Q_{r_M-1}(c, b), Q_{r_M+1}(c, b)) \tag{6.16}$$

then the minimum value of $Q_r(c, b)$ (with respect to r) is given by

$$\min(Q_{r_M}(c, b), Q_0(c, b)) \ . \tag{6.17}$$

b) If no such r_M exists then the minimum of $Q_r(c, b)$ is given by

$$\begin{aligned} & \min(Q_0(c, b), Q_p(c, b)) \\ = \ & \min(V_{p+1}(c), U_p(b)) \ . \end{aligned} \tag{6.18}$$

Remark

In practice, r_M may be guessed (and checked) in order to find (6.17). Alternatively, it may be guessed that no such r_M exists, which may be checked by finding that either

$$Q_1(c,b) < Q_2(c,b)$$

or

$$Q_{p-1}(c,b) > Q_p(c,b) \ .$$

Choice of c and b

Suppose that an experimenter wishes to construct both bounds (6.5) and (6.6) for comparison with the control and the best, with specified simultaneous confidence P^*. Then P^* is equal to the right hand side of (6.7) and, for particular $F_0(\cdot)$ and $F_1(\cdot)$ distributions, appropriate values of c and b, say c^* and b^*, may be calculated. The values of c^* and b^* satisfying P^* are not unique, however, since b^* may be decreased at the expense of increasing c^* and vice versa. With a smaller value of b^*, more sensitive comparisons with the best are achieved but comparisons with the control are less sensitive. This may be useful if an experimenter is reasonably sure that some of the populations are better than the control. Similarly, a smaller value of c^* is appropriate if the bounds (6.5) (for comparison with the control) are required to be more sensitive than the intervals (6.6) (for comparison with the best).

Three cases of special interest were considered by Bofinger and Mengersen:

(i) Comparisons with the control are required to be as sensitive as possible. In this case, the minimum possible value of c^* is taken, with

$$V_{p+1}(c^*) = P^* \ .$$

(ii) Comparisons with the best are required to be as sensitive as possible. In this case, the minimum possibe value of b^* is taken, with

$$U_p(b^*) = P^* \ .$$

(iii) Comparisons with the control and with the best are required to be equally sensitive. In this case, $b^* = c^*$.

Case (i) The value c_C is defined so that

$$Q_0(c_C, b) = V_{p+1}(c_C) = P^* \tag{6.19}$$

and hence b_C is found to satisfy

$$\min_{1 \le r \le p} Q_r(c_C, b_C) = P^* \ .$$

Case (ii) The value b_B is defined so that

$$Q_p(c, b_B) = U_p(b_B) = P^* \tag{6.20}$$

and hence c_B is found to satisfy

$$\min_{0 \le r \le p-1} Q_r(c_B, b_B) = P^* \ . \tag{6.21}$$

Case (iii) The value c_{CB} is required to satisfy

$$\min_{0 \le r \le p} (c_{CB}, c_{CB}) = P^* \ . \tag{6.22}$$

In order to solve (6.19), Theorem 6.2 and the remarks following it may be useful. In order to solve (6.20), Bofinger and Mengersen's Theorem 2.3 may be used, which is given below in Theorem 6.3. Using this theorem, we can take as c_B the value satisfying (6.20) for $r = p - 1$, if this value is larger than that satisfying (6.20) for $r = p - 2$ and $r = 0$. Similarly, (6.21) may be solved with the aid of their Theorems 2.4 and 2.5, which are given below as Theorems 6.4 and 6.5. (Again, the proofs for the following theorems may be found in the paper by Bofinger and Mengersen (1987).)

Theorem 6.3

If c_C is defined by (6.19) and if $c_{(B)}$ satisfies

$$Q_{p-1}(c_{(B)}, b_B) = P^* \tag{6.23}$$

then

$$c_B = \max(c_{(B)}, c_C) \ . \tag{6.24}$$

Theorem 6.4

a) If $\exists c_M$ and $r_M \in [2, p-1]$ such that

$$P^* = Q_{r_M}(c_M, c_M) \leq \min\left(Q_{r_M-1}(c_M, c_M), Q_{r_M+1}(c_M, c_M)\right)$$

then

$$c_{CB} = \max(c_M, c_C)$$

where c_C is defined by (6.19).

b) If no such r_M and c_M exist then

$$c_{CB} = \max(c_C, b_B)$$

with b_B defined by (6.20).

Remark

By Theorem 6.4, we can "guess" a value of r between 2 and $p-1$, find the corresponding value of c_M and test if this is less than the value of c_M for $r-1$ and $r+1$. Notice, however, that if c_M and r_M exist but $c_M < c_C$, it would be more efficient to use case (i) instead of case (iii). That is, instead of requiring equally sensitive comparisons with the control and the best, the experimenter should obtain the more sensitive comparisons with the best. Similarly, if c_M and r_M do not exist and if $c_C > b_B$, case (i) should again be used, while if $c_C < b_B$ case (ii) should be applied.

Bofinger and Mengersen's Theorem 2.5 is useful for the case in which $F_0(\cdot) = F_1(\cdot)$ (so that only location parameters differ between the control and treatment distributions). This theorem is given below in Theorem 6.5 (with proof omitted).

Theorem 6.5

If $F_0(\cdot) = F_1(\cdot)$, then the required value of $c = b$ satisfies

$$\min_{0 \leq r \leq p} Q_r(c, c) = \min(Q_0(c, c), Q_{[\frac{p+1}{2}]}(c, c)) \ .$$

6.3 Selection Decisions

The bounds (6.5) and (6.6) may be used for selection decisions. With joint confidence given by (6.7), the following assertions may be made:

a) All populations $\pi_i, i \in G$, are demonstrably better than the control population π_0.

b) If $R(p) \in G$ (that is, G is not empty) *and*

$$X_{R(p)} \geq \max_{j \neq 0, R(p)} X_j + b$$

then we can assert (as well as (a)) that $R(p) = \rho(p)$. That is, $\pi_{R(p)}$ is better than all of the other populations, including the control. Otherwise (for G non-empty) we can assert that $\pi_{R(p)}$ is better than π_0 and that $\theta_{R(p)}$ is within

$$\max_{j \neq 0, R(p)} X_j + b - X_{R(p)}$$

of $\theta_{\rho(p)}$. If this difference is sufficiently small, the experimenter may still be interested in $\pi_{R(p)}$.

c) If some other population $\pi_i, i \in G$ is of interest, then an impression of its performance compared with the best population may be gained by examining the value

$$X_i - X_{R(p)} - b \ .$$

If this value is not too small, the population may still be worthy of consideration due to other factors such as cost efficiency or availability. Notice, however, that it is the bounds that are provided to the experimenter, and it is the experimenter who may make appropriate selection decisions based on this information.

d) If G is empty (so that no population is found to be demonstrably better than the control), then again the bounds (6.5) are useful in considering the magnitude of the differences between the populations and the control.

e) If the experimenter wishes to select a subset containing all populations which are demonstrably better than the control *and* not demonstrably worse than the best population, the subset would include all π_i for which

$$X_i \geq \max(X_0 + c, X_{R(p)} - b) \ .$$

That is, all populations with $L_i = 0$ *and* positive upper bound given by I_i would be selected.

Bofinger and Mengersen's method may be favourably compared with the Indifference Zone-type approaches proposed by Turnbull (1976), Bechhofer and Turnbull (1978) and Bristol and Desu (1985) with respect to the following:

- The control may be unknown and may have a different distribution from that of the other experimental populations. The three alternative procedures above require a fixed (known) standard.
- The five constants which must be specified by the experimenter under the Indifference Zone-type procedures, which define the indifference zones and probabilities of correct selection, are not required under the confidence bound approach. Instead, it is necessary to prescribe P^* and which of the three cases is of particular interest.
- Specifying the confidence bounds leaves the onus of selection on the experimenter. More information about the comparative performance of the populations with respect to the control and to the best is available to the experimenter under the confidence bound approach than merely the selection statements provided by the Indifference Zone-type approaches.

6.4 Special Results for the Normal Means Case with Common Unknown Variance

Let X_i be the sample mean based on n_1 independent observations from π_i, distributed normally with unknown variance σ^2, $i = 1,..,p$. Let X_0 be the sample mean based on n_0 independent observations from the control population π_0, with the same (unknown) variance σ^2. Then, with $\Phi(\cdot)$ denoting the standard normal distribution,

$$F_0(x - \theta_0) = \Phi\left((x - \theta_0)n_0^{1/2}/\sigma\right)$$

and

$$F_1(x-\theta_i) = \Phi\left((x-\theta_i)n_1^{1/2}/\sigma\right), \quad i = 1,..,p \ .$$

The variance σ^2 is estimated by the usual S^2, such that $\nu S^2/\sigma^2$ has a chi-squared distribution on ν degrees of freedom.

Under this formulation, we have:

$$G = \{i : X_i - X_0 > cS(n_0^{-1} + n_1^{-1})^{1/2}\} \tag{6.25}$$

$$L_i = \min(0, X_i - X_0 - cS(n_0^{-1} + n_1^{-1})^{1/2}) \tag{6.26}$$

and

$$\begin{aligned} I_i = \ & [\min(0, X_i - \max_{j\neq i,0} X_j - bS2^{1/2}n_1^{-1/2}), \\ & \max(0, X_i - \max_{j\in G,\neq i} X_j + bS2^{1/2}n_1^{-1/2})] \ . \end{aligned} \tag{6.27}$$

The joint confidence of the bounds (6.26) and (6.27) obtained by Bofinger and Mengersen is given in Theorem 6.6.

Theorem 6.6

With G, L_i and I_i defined by (6.25), (6.26) and (6.27) respectively, we have

$$\begin{aligned} & \inf_{\underline{\theta}\in\Omega} P\{\theta_i - \theta_0 > L_i \ \ \forall i = 1,..,p \ \text{ and } \ \theta_i - \max_{j\neq i,0}\theta_j \in I_i \ \ \forall i \in G\} \\ = \ & \min_{0\leq r\leq p} E_S\{P_{p-r+1}(cS/\sigma, \rho), P_r(bS/\sigma, 1/2)\} \end{aligned} \tag{6.28}$$

where $E_S\{\cdot\}$ indicates expectation over the distribution of S,

$$\rho = (1 + n_0/n_1)^{-1} \tag{6.29}$$

and

$$P_{p-r+1}(y,\rho) = \int \left(\Phi(\frac{u\rho^{1/2} + y}{(1-\rho)^{1/2}})\right)^{p-r} d\Phi(u) \ . \tag{6.30}$$

Proof

Let

$$\Omega_r = \{\underline{\theta} : \#\gamma = r\} \ .$$

Then, conditionally on S, the left hand side of (6.28), with the infimum over $\underline{\theta} \in \Omega_r$ instead of all Ω, is shown to be equal to

$$V_{p-r+1}(cS(n_0^{-1} + n_1^{-1})^{1/2})U_r(bS2^{1/2}n_1^{-1/2})$$

where

$$\begin{aligned} & V_{p-r+1}(cS(n_0^{-1} + n_1^{-1})^{1/2}) \\ = & \int (\Phi((x + cS(n_0^{-1} + n_1^{-1})^{1/2})n_1^{1/2}/\sigma))^{p-r} d\Phi(xn_0^{1/2}/\sigma) \\ = & P_{p-r+1}(cS/\sigma, \rho) \ . \end{aligned}$$

When $n_0 = n_1$, this last expression reduces to

$$U_r(bS2^{1/2}n_1^{-1/2}) = P_r(bS/\sigma, 1/2) \ .$$

The result follows after taking the expected value over the distribution of S and the infimum over

$$\Omega = \bigcup_{r=0}^{p} \Omega_r \ .$$

Lemma 6.1 and Theorems 6.2 and 6.3 hold in this special case if $Q_r(c, b)$ is replaced by

$$Z_r(c, b) = E_S\{P_{p-r+1}(cS/\sigma, \rho)P_r(bS/\sigma, 1/2)\} \ .$$

Choice of c and b Following the general results obtained in Section 6.2, Bofinger and Mengersen considered the three cases of interest for the normal distribution. It is assumed that n_0, n_1 and the joint confidence P^* have been specified by the experimenter, as well as the particular case of interest. Also, in most practical situations, more observations will be taken from the control population than from the experimental populations, so Bofinger and Mengersen considered only the situation in which $n_0 \geq n_1$ (and hence $\rho \leq 1/2$).

Case (i) Comparisons with the control are required to be as sensitive as possible.

From Section 6.2, the minimum possible value of c, say c_C, is given by

$$E_S\{P_{p+1}(c_C S/\sigma, \rho)\} = P^* \tag{6.31}$$

and b_C is calculated such that (6.28) equals P^*. The remarks in Section 6.2 regarding computation of c_C and b_C are applicable for this special case. As discussed in Chapter 4, values of c_C have been tabulated by a number of authors. Using two-dimensional quadrature routines from the NAG Fortran library, with the outer integral computed using the 64-point Gauss-Hermite quadrature routine D01BAF and the inner integral computed using the automatic adaptive routine D01AMF, the values of $\{c_C, b_C\}$ were computed for various $\{P^*, \rho, \nu, p\}$ combinations. These values are presented in Table B.5 (located in Appendix B). Beginning with "guessed" values of c_C and b_C, iteration was continued until the probability (6.28) was within 1.0E-5 of P^*, for each particular $\{P^*, \rho, \nu, p\}$ set.

Case (ii) Comparisons with the best are required to be as sensitive as possible.

In this case, the minimum possible value of b, say b_B, is to be used and is given by

$$E_S\{P_p(b_B S/\sigma, 1/2)\} = P^* \ .$$

Then c_B is the value such that (6.28) equals P^*. From the results in Section 6.2, c_B is given by

$$c_B = \max(c_{(B)}, c_C) \tag{6.32}$$

with c_C given by (6.31) and $c_{(B)}$ satisfying

$$Z_{p-1}(c_{(B)}, b_B) = P^* \ .$$

The values of b_B may be found from c_C values in Table B.5 with $\rho = 0.5$, by taking a value of p one less than that for c_C. That is,

$$b_{B,p} = c_{C,p-1,0.5}$$

where the subscripts p and $p-1$ denote the number of experimental populations and 0.5 is the value of ρ.

The values of $c_{(B)}$, which are independent of ρ, were computed in the same way as described for Case (i) and are presented in Table B.6 (in Appendix B). Appropriate values of c_B may be found using Table B.5 and Table B.6 with (6.32), for $\rho = 0.1(0.1)0.5$. For $\rho > 0.5$, the values of c_C (and hence c_B) may be found from

Gupta, Panchapakesan and Sohn's (1985) tables. Krishnaiah and Armitage (1966) give some values of c_C for $\rho = 0$.

Case (iii) Comparisons with the control and with the best are required to be equally sensitive.

With

$$c_{CB} = b_{CB} 2^{1/2} (1 - \rho)^{1/2} ,$$

values of c_{CB} are computed as the solution to

$$P^* = \min_{0 \leq r \leq p} Z_r(c_{CB}, b_{CB}) .$$

For various $\{P^*, \rho, \nu, p\}$ combinations, these c_{CB} values are presented in Table B.7 (in Appendix B). Notice, however, that some of the values in this table have been replaced by the notation (i) or (ii). This indicates that for this particular $\{P^*, \rho, \nu, p\}$ combination, it is more efficient to use one of the other two cases. For example, when

$$b_{CB} = b_B$$

it is preferable to use Case (ii), since a smaller value of c may be used with b_B with the same confidence. When this occurs, the value in Table B.7 is replaced by the notation (ii). Similarly, when

$$c_{CB} = c_C$$

the value in Table B.7 is replaced by the notation (i), indicating that it is preferable to use Case (i), since a smaller value of b may be used with c_C with the same confidence.

Where possible, the values in Tables B.5, B.6 and B.7 were compared with values in other tables, such as those of Gupta, Panchapakesan and Sohn (1985). Response surfaces described in Chapter 4 were fitted to a number of sections of the tables and outliers (identified by large RMS values) were verified by computing them separately and comparing different values of r in detail.

6.4.1 Upper bound on the joint confidence

Bofinger and Mengersen noticed that

$$Z_r(c,b) \geq E_S\{P_{p-r+1}(cS/\sigma,\rho)\}E_S\{P_r(bS/\sigma,1/2)\} \tag{6.33}$$

where

$$E_S\{P_{p-r+1}(cS/\sigma,\rho)\}$$

is the probability that $(p-r)$ multivariate Student t variables, with degrees of freedom ν and common correlation ρ, are all less than c. This expression is straightforward to compute and has been tabulated by a number of authors, as discussed in Chapter 4. It may be used to find useful upper bounds for b_C, c_B and c_{CB}. For Cases (i) and (ii), Table 6.1 presents values of b_C and c_B (depending on the case) such that expression (6.33) equals 0.75,0.95 for various ρ, ν and p. (Case (iii) was omitted because Cases (i) and (ii) are preferable for many of the values.) The superscript accompanying each value is the amount by which the upper bound exceeds the corresponding value in Tables B.5 and B.6 (as appropriate).

In Case (i), the r value which minimises (6.28) was also found to minimise (6.33) for most $\{P^*, p, \rho, \nu\}$ combinations, but for larger ν, ρ and p the required r value for (6.33) may exceed that for (6.28). (All r values tend to fall between 2 and 5.) The bounds become more satisfactory for small p and larger ν.

For case (ii), the r value which minimises (6.33) was also found to minimise (6.28). For large ν these bounds are not too conservative.

6.5 Related Problems

6.5.1 Simultaneous comparison with a control and with the t best

A generalisation of the results of this chapter may be made by considering the problem of simultaneously comparing populations with a control and with the t best $(0 \leq t \leq p)$. This may be useful in practice for a number of reasons. For example,

Table 6.1: Upper bounds for various combinations of $\{P^*, p, \rho, \nu\}$

Upper bounds for b_C for case (i).

		$p = 4$		$p = 9$		$p = 16$		$p = 20$	
ν	ρ	0.75	0.95	0.75	0.95	0.75	0.95	0.75	0.95
10	0.10	$1.28^{.02}$	$2.35^{.01}$	$1.86^{.07}$	$2.88^{.05}$	$2.27^{.14}$	$3.28^{.11}$	$2.43^{.17}$	$3.43^{.13}$
	0.30	$1.31^{.01}$	$2.36^{.01}$	$1.95^{.08}$	$2.94^{.06}$	$2.37^{.13}$	$3.35^{.10}$	$2.53^{.16}$	$2.51^{.12}$
	0.50	$1.36^{.01}$	$2.39^{.01}$	$2.06^{.08}$	$3.01^{.06}$	$2.49^{.12}$	$3.43^{.08}$	$2.66^{.15}$	$3.62^{.12}$
40	0.10	$1.22^{.01}$	$2.12^{.00}$	$1.71^{.02}$	$2.51^{.01}$	$2.02^{.03}$	$2.76^{.02}$	$2.14^{.04}$	$2.86^{.02}$
	0.30	$1.25^{.01}$	$2.12^{.00}$	$1.78^{.02}$	$2.53^{.00}$	$2.11^{.03}$	$2.81^{.02}$	$2.23^{.03}$	$2.90^{.01}$
	0.50	$1.29^{.00}$	$2.15^{.01}$	$1.87^{.02}$	$2.58^{.00}$	$2.21^{.02}$	$2.86^{.01}$	$2.33^{.03}$	$2.97^{.02}$
120	0.10	$1.20^{.00}$	$2.09^{.01}$	$1.68^{.01}$	$2.43^{.00}$	$1.97^{.01}$	$2.67^{.01}$	$2.08^{.01}$	$2.75^{.00}$
	0.30	$1.23^{.00}$	$2.09^{.01}$	$1.74^{.00}$	$2.46^{.00}$	$2.05^{.00}$	$2.70^{.00}$	$2.17^{.01}$	$2.79^{.00}$
	0.50	$1.28^{.00}$	$2.10^{.00}$	$1.83^{.00}$	$2.50^{.00}$	$2.15^{.00}$	$2.75^{.00}$	$2.26^{.00}$	$2.85^{.00}$

Upper bounds for c_B for case (ii).

	$p = 4$		$p = 9$		$p = 16$		$p = 20$	
ν	0.75	0.95	0.75	0.95	0.75	0.95	0.75	0.95
10	$1.63^{.04}$	$2.59^{.04}$	$2.34^{.11}$	$3.28^{.09}$	$2.78^{.16}$	$3.75^{.14}$	$2.94^{.18}$	$3.93^{.15}$
40	$1.51^{.01}$	$2.28^{.01}$	$2.08^{.02}$	$2.74^{.01}$	$2.40^{.03}$	$3.02^{.01}$	$2.52^{.04}$	$3.13^{.01}$
120	$1.49^{.00}$	$2.22^{.01}$	$2.03^{.00}$	$2.64^{.00}$	$2.33^{.00}$	$2.90^{.01}$	$2.44^{.01}$	$2.99^{.00}$

an experimenter may indeed wish to identify more than one good population, with confidence that all of those identified perform better than some control or standard. In another situation, more than one good population may be identified and, from these, a single population may be finally selected on some other basis, such as cost or availability.

In this case, the θ values are assumed to be ordered as follows:

$$\theta_{\rho(1)} \leq \theta_{\rho(2)} \leq ... \leq \theta_{\rho(p-t)} < \theta_{\rho(p-t+1)} \leq ... \leq \theta_{\rho(p)} \ .$$

Those populations which are truly better than the control are included in the set γ, defined by (6.3) and those which are observed to be demontrably better than the control are included in the set G, defined by (6.4). The t populations (excluding the control) which are truly best are those $\pi_i, i \in \gamma_t$, where

$$\gamma_t = \{\rho(k-t+1), ..., \rho(k)\} \ .$$

The lower bounds L_i on $\theta_i - \theta_0 \ \forall i = 1, .., p$, given by (6.5), are again constructed. Provided $\#G \geq t$, intervals are considered for

$$\theta_i - \theta_{\rho(p-t+1)}$$

$\forall i \in G$ of

$$I_{t_i} = \{\min(0, X_i - X_{R(p-t+1|i)} - b), \max(0, X_i - X_{R(p-t+1|i),G} + b\} \qquad (6.34)$$

where $X_{R(p-t+1|i)}$ is the tth largest X_j, $j = 1, .., p, \neq i$ and $X_{R(p-t+1|i),G}$ is the tth largest X_j, $j \in G, \neq i$.

When $\#G < t$, no upper or lower bound for the interval I_{t_i} is defined. When $\#G = t$, only the lower bound is defined.

Theorem 6.7

With G, L_i and I_{t_i} defined by (6.4), (6.5) and (6.34) respectively, we have

$$\begin{aligned} & \inf_{\underline{\theta} \in \Omega} P\{\theta_i - \theta_0 > L_i \ \forall i = 1, .., p \ \text{ and } \ \theta_i - \theta_{\rho(p-t+1)} \in I_{t_i} \ \forall i \in G\} \\ = \ & \min_{t \leq r \leq p} \{V_{p-r+1}(c) U_{r,t}(b)\} \end{aligned} \qquad (6.35)$$

where

$$U_{r,t}(b) = 1,\ r = 0,..,t-1\ ;$$

$$U_{r,t}(b) = t\int F_1^{r-t}(x+b)(1-F_1(x))^{t-1}dF_1(x)\ ,\ t \le r \le p$$

and

$$V_j(c) = \int (F_1(x+c))^{j-1}dF_0(x)\ .$$

Proof

Replace E_b in the proof of Theorem 6.1 by

$$E_b' = \{X_j - \theta_j < X_i - \theta_i + b\ \forall i \in \gamma_t, j \in \gamma \cap \overline{\gamma}_t\}$$

and let $g = \#\gamma$.

Then

$$P\{E_b'\} = U_{r,t}(b)$$

and the event E_c is defined in (6.9) with

$$P\{E_c\} = V_{p-g+1}(c)\ .$$

By a direct extension of the proof of Theorem 6.1, then, the result follows.

As detailed earlier, selection decisions may be made based on the bounds L_i and I_{t_i}. With joint confidence given by (6.35), we may make the following statements:

a) All populations $\pi_i, i \in G$ are demonstrably better than the control population π_0.

b) If $R(p-t+1) \in G$ *and* $X_{R(p-t+1)} \ge \max_{j \notin G, \neq 0} X_j + b$, then we can assert both (a) *and* that $\{\rho(p-t+1),..,\rho(p)\} = \{R(p-t+1),..,R(p)\}$, that is, that $\pi_{R(p-t+1)},..,\pi_{R(p)}$ are indeed better than all the other populations, including the control.

c) For those $\pi_i, i \in G$, the value

$$X_i - X_{R(p-t+1)} - b$$

gives some information about the performance of π_i compared with the tth best population. The population may still be worthy of the attention of the experimenter if this value is not too far below zero..

d) If $\#G < t$, so that there are not t populations which are demonstrably better than the control, the experimenter may wish to reduce the value of t or the joint confidence. In any case, the performance of the experimental populations with respect to the control may be assessed using the numerical comparisons L_i.

e) A subset of populations which are demonstrably better than the control and not demonstrably worse than the t best populations includes all π_i for which

$$X_i \geq \max(X_0 + c, X_{R(p-t+1)} - b) \ .$$

For the normal means case with common unknown variance, the joint confidence is given by

$$Z_{r,t}(c,b) = \min_{t \leq r \leq p} E_S\{P_{p-r+1,1}(cS/\sigma, \rho), P_{r,t}(bS/\sigma)\} \ .$$

A lower bound on this joint confidence is given by

$$Z_{r,t}(c,b) \geq E_S\{P_{p-r+1,1}(cS/\sigma, \rho)\} + E_S\{P_{r,t}(bS/\sigma)\} - 1 \ .$$

Conservative values of c and b statisfying this lower bound for the normal means case may be found by interpolation in tables such as Gupta, Panchapakesan and Sohn (1985) (for c values) and Mengersen and Bofinger (1987) (for b values; see Table B.1). Satisfactory interpolation methods for the latter tables are described in Chapter 4.

6.5.2 Simultaneous comparisons with restrictions on the "best"

The approach taken in this chapter may be used to develop solutions to other problems. For example, Chen (1985) uses an Indifference Zone approach to select the

best normal population provided that its mean does not lie in a prespecified "undesirable" region of the indifference zone (as discussed in Chapter 2). By the nature of the selection procedure, it is also possible that no population is selected if $X_{R(p)}$ and $X_{R(p-1)}$ differ by too little. That is, even if $\theta_{\rho(p)}$ lies outside the "undesirable" region, no selection will be made if $\pi_{\rho(p)}$ is not distinctly best.

An alternative approach may be for the experimenter to specify a minimum value θ_0 such that, if θ_i from any population π_i is less than θ_0, that population should not be selected. (This corresponds to Chen's "undesirable" region.) This value θ_0 may be used as a known standard in the procedure described in this chapter. In this way, with a specified joint confidence, only those populations which are demonstrably better than the standard are considered *and* all those populations which are not demonstrably worse than the best are identified. The confidence bound approach gives numerical information about the performance of the populations compared with the control and with the best, rather than only the selection decisions obtained under Chen's procedure.

6.6 Example

Consider the example given in Chapter 3, in which 21 first year students each judge 8 lecturers on the basis of a number of specified attributes. After early exclusion of two of the lecturers, the remaining six lecturers are to be compared in order to select the best lecturer. Suppose, however, that it is decided to award the prize only if the chosen lecturer is shown to be demonstrably better than the lecturer who was awarded the prize in the previous year. That is, the students wish to select the best lecturer, provided that he or she is better than a control (the previous year's winner). Suppose that the control lecturer in this case is Wendy. (Notice that the students are taking Wendy to be a control, rather than taking her previous year's score to be a (known) standard. This is more appealing because, for example, different students are involved in the judging procedure this year.)

From Chapter 3, the observed final scores for each of the lecturers are as follows:

Carole	Debbie	Wayne	Wendy	Russell	Janice
134.7	125.2	118.7	111.5	89.5	70.2

with $s^2 = 107.04$ on 140 degrees of freedom.

In this case, $p = 5$, $s(n_0^{-1} + n_1^{-1})^{1/2} = 3.19$ and $\rho = 0.5$.

Consider the following cases of interest:

i) Comparisons with Wendy are required to be as sensitive as possible.

ii) Comparisons with the best lecturer (other than Wendy) are required to be as sensitive as possible.

iii) Comparisons with Wendy and the best (other) lecturer are required to be equally sensitive.

Suppose that the students decide to investigate the decisions that can be made at both the 75% and 95% confidence levels. The appropriate critical values found from Tables B.5, B.6 and B.7 for each of these cases are as follows:

Case		$P^* = .75$	$P^* = .95$
(i)	c_C	1.40	2.26
	b_C	1.45	2.21
(ii)	$c_{(B)}$	1.66	2.34
	b_B	1.31	2.18
(iii)		use case (i)	

In this situation, it is more useful to use Case (i) rather than consider equally sensitive comparisons with the control and the best.

Lower confidence bounds on $\theta_i - \theta_0$, where θ_0 is the true score for Wendy and θ_i is the true score for the other lecturers C_i, $i = 1, .., 5$, are constructed as follows:

Case	P^*	Carole	Debbie	Wayne	Russell	Janice
(i)	0.75	0	0	0	-26.47	-45.77
	0.95	0	0	-0.01	-29.21	-48.51
(ii)	0.75	0	0	0	-27.30	-46.60
	0.95	0	0	-0.26	-29.46	-48.77

Hence, for both Cases (i) and (ii), at the 75% confidence level, the students would select Carole, Debbie and Wayne as demonstrably better than Wendy. At the 95% level, however, they would only include in this subset Carole and Debbie, although by investigating the bounds, they may consider including Wayne for case (i). (Notice that this could not be considered under the subset selection or indifference zone approaches.)

In order to compare these lecturers with the best, intervals I_i were constructed:

Case	P^*	Carole	Debbie	Wayne
(i)	0.75	(0,14.13)	(-14.13,0)	(-20.63,0)
	0.95	(0,16.55)	(-16.55,0)	
(ii)	0.75	(0,13.68)	(-13.68,0)	(-20.18,0)
	0.95	(0,16.45)	(-16.45,0)	

Hence, for both P^* values and for all three cases, Carole may be asserted to be both demonstrably best *and* demonstrably better than Wendy.

Chapter 7

ROBUSTNESS TO NORMALITY OF A SELECTION RULE

7.1 Introduction

In the previous three chapters, a special case on which much emphasis has been placed is that of independent normal distributions, with common unknown variance and equal numbers of observations from each experimental population. For this case, particular results were derived and tables of percentage points satisfying probability requirements were presented. A question that now arises is how robust the procedures are to deviations from the assumptions on which this special case is based.

As discussed in Chapters 2 and 4, a number of authors have considered different correlation structures for the problems of selecting the one best population (general degrees of freedom) and selecting t best populations (infinite degrees of freedom). For example, Gupta and Panchapakesan Sohn (1985) have produced extensive tables of percentage points satisfying

$$P_{k,1}^{(\nu)}(d,\rho) = \int\int (1 - F(\frac{u\rho^{1/2} - dS/\sigma}{(1-\rho)^{1/2}}))dF(u))^{k-1} dG_{\nu}(S/\sigma)$$

for different correlation values ρ. Some indication of the effect of non-independence on the required percentage point or resulting probability of correct selection may be gained by examining these tables.

If a common variance cannot be assumed, modifications of the methods described in the previous chapters may be made. These may take the form of two-stage procedures, such as those proposed by Dudewicz and Dalal (1975), Rinott (1978) and Santner and Tamhane (1985) for selecting the best population, or Bofinger and Lewis' (1987) (confidence bound) approach for comparing populations with a control and with the best.

The problem of unequal numbers of observations may similarly require a modification of the probability expressions. Results have been obtained for the case $t = 1$, some of which were discussed in Chapter 2 and which were detailed by Gupta and Panchapakesan (1979, pp.22-25, 233-237, 244-245, 417-420). For $t = 1$ under the confidence bound approach, Hsu (1984) has proposed a methodology for unbalanced designs which follows that established for balanced designs, with the exception that a vector of d values is used instead of a single value. This was also detailed in Chapter 2.

Although Dunnett (1980a,1980b,1982),Tamhane (1979) and others have investigated the effects of non-normality on procedures for simultaneous comparisons, the robustness of ranking and selection procedures to non-normality has only recently been considered. As discussed in Chapter 2, Dudewicz and Mishra (1985) compared the performance of Bechhofer's (1984) procedure for selecting the t best normal means (with known variance) under a number of non-normal distributions. Exact results were given for the uniform distribution and simulated results for the t-distribution on 3 degrees of freedom. Across a range of (k, t) combinations and for two parameter configurations, the authors established that Bechhofer's procedure is indeed quite robust to such deviations from normality. These results beg further questions, however, such as how skewness affects the procedure and whether the subset selection and confidence bound approaches enjoy such robustness.

In this chapter, robustness to non-normality is considered in detail for the procedure described in Chapter 5, which aims to select from k populations a subset containing the t populations with the largest location parameters. Using a method suggested by Gupta and Sohn (1985), Tukey's generalised lambda distribution (GLD) is used to approximate various continuous distributions and to evaluate the probability of correct selection (PCS) arising under these distributions. This is achieved by considering the performance of the lower bound on the PCS that was obtained in Chapter 5. If the bound performs satisfactorily, the PCS will also perform satisfactorily and hence the procedure may be asserted to be robust for that particular case. If, however, the bound performs poorly, this is not sufficient indication that the PCS is also unsatisfactorily small (since it may be only that the bound is extremely conservative in these cases). Remedies to this problem are suggested.

This chapter indicates preliminary investigations into the question of non-normality and is not intended to answer all of the issues that arise from these considerations. Where appropriate, there will be mention of these issues and of further investigations that are at present being undertaken.

The next section details the particular selection problem to be considered in this chapter. The generalised lambda distribution is investigated in Section 7.3 and is used in Section 7.4 to compute lower bounds for the probability of correct selection under the selection procedure. A Monte Carlo experiment is also undertaken to investigate the true probability of correct selection for cases in which the lower bound is unsatisfactory. A more general approach to the robustness problem is considered in Section 7.5 and an example of this approach is discussed in Section 7.6.

7.2 The Selection Problem

Consider the problem of selecting the t best of k populations, based on n independent observations from each population. The selection goal investigated in Chapter 5 is:

Goal G: Select $\pi_i, \forall i \in \gamma_t$, where

$$\gamma_t = \{\rho(k-t+1), \rho(k-t+2), .., \rho(k)\} \ .$$

The procedure proposed in Chapter 5 to satisfy this goal is:
Procedure P: Select $\pi_i, \forall i \in G(d)$, where

$$G(d) = \{i : X_i > X_{R(k-t+1)} - d\sigma\} \tag{7.1}$$

with d satisfying

$$P_L = P^* \tag{7.2}$$

where

$$P_L = (k-t+1)P_{k-1,t-1}(d) - (k-t)P_{k,t-1}(d) + P_{k-t+1,1}(d) - 1 \ , \tag{7.3}$$

P^* is the prespecified probability of correct selection and $P_{k,t}(d)$ is given by

$$P_{k,t}(d) = t \int (F(x+d))^{k-t}(1-F(x))^{t-1} dF(x) \ . \tag{7.4}$$

It is assumed that independence exists between all observations, and that all k populations have the same distribution with common known variance $\sigma^2 = 1$, but it is *not* assumed that the form of the distribution is normal.

Using the percentage point satisfying (7.2), with $F(\cdot)$ taken to be the normal distribution, we wish to examine the value of P^* using non-normal $F(\cdot)$. That is, if Procedure P were used with d values tabulated in Table B.3 (in Appendix B), we wish to examine the effect on the resulting probability of correct selection if the underlying distribution is in fact non-normal.

In order to investigate this question, use of the GLD was investigated. This distribution, introduced by Tukey in 1960 for symmetric distributions and generalised by Ramberg and Schmeisser (1972,1974) to include asymmetric distributions, was used originally to generate random variables. Recently, as discussed in Chapter 2, Gupta and Sohn (1985) have used this distribution to develop and evaluate a selection procedure based on sample medians. The latter authors showed that satisfactory approximations to a very wide class of distributions may be obtained

using the GLD, by using different combinations of the distribution's parameters. Gupta and Sohn also showed that the GLD was useful in obtaining approximations to the percentage points of (7.4) for $t = 1$ and suggested that the GLD could be used to test robustness of selection procedures to non-normality.

7.3 The Generalised Lambda Distribution

The GLD may be expressed as follows:

$$x = F^{-1}(\mathrm{p}) = \lambda_1 + \lambda_2^{-1}(p^{\lambda_3} - (1 - p^{\lambda_4})) \tag{7.5}$$

where $F^{-1}(\cdot)$ is the inverse of the distribution function $F(\cdot)$ of the GLD, p is a uniform (0,1) random variable and where λ_1 and λ_2 are location and scale parameters respectively, and λ_3 and λ_4 are measures of skewness and kurtosis. (Notice that, although λ_2 is described as a scale parameter, it is not necessarily positive.)

For $\lambda_1 = 0$, Ramberg and Schmeiser showed that the even moments of the GLD, when they exist, are given by

$$E(X^k) = \lambda_2^{-k} \sum_{i=0}^{k} \binom{k}{i} (-1)^i \beta(\lambda_3(k-i) + 1, \lambda_4 i + 1)$$

where β denotes the beta function. The kth moment exists if and only if $-1/k < \min(\lambda_3, \lambda_4)$ (so that the arguments of the beta function are nonnegative).

Ramberg and Schmeisser (1974) identified the four regions in which the GLD is a legitimate probability distribution, that is, where the density function $f(x)$ is nonnegative for all x and $\int_{-\infty}^{\infty} f(x)dx = 1$. For symmetric distributions, λ_3 is equal to λ_4 and the sign corresponding to λ_3 must be the same as the sign corresponding to λ_2. For these distributions, all odd central moments are zero and both the mean and median equal λ_1. The standardised second and fourth central moments (σ^2, α_4) are given by

$$\sigma^2 = 2(1/(2\lambda_3 + 1) - \beta(\lambda_3 + 1, \lambda_3 + 1))/\lambda_2^2 ,$$

$$\alpha_4 = \frac{1/(4\lambda_3 + 1) - 4\beta(\lambda_3 + 1, 3\lambda_3 + 1) + 3\beta(2\lambda_3 + 1, 2\lambda_3 + 1)}{2[1/(2\lambda_2 + 1) - \beta(\lambda_3 + 1, \lambda_3 + 1)]^2} .$$

7.3.1 Generating random variables

Many symmetric and asymmetric distributions may be approximated in the form of the GLD. Given the first four moments (in terms of the parameters of the known distribution), Ramberg and Schmeisser (1972,1974) illustrated how the appropriate parameter combinations $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ for the GLD which approximates the particular distribution may be obtained. If the first four moments do not exist $(\lambda_3 \leq -0.25)$, the parameter combinations may be obtained by using the percentile points of the distribution. Table 7.1 lists combinations of parameters for the GLD, approximating a number of common symmetric and asymmetric distributions, which were tabulated by Ramberg and Schmeisser.

For example, for the normal distribution with mean θ and variance σ^2, $\alpha_4 = 3.0$ and, from Table 7.1, $(\lambda_1, \lambda_2, \lambda_3) = (\theta, 0.1975/\sigma, 0.135)$. Hence, the inverse GLD which approximates the inverse normal distribution is given by

$$x = \theta + \sigma(p^{0.135} - (1-p)^{0.135})/0.1975 \ , \tag{7.6}$$

$0 < p < 1$. This may be used to generate approximately normal random variables.

7.3.2 Approximating empirical distributions

Ramberg and Schmeisser suggested that empirical distributions may be approximated by using estimates of the moments based on the random sample X_i, $i = 1, .., n$. For a wide range of distributions, they generated random variables, obtained estimates of the parameter values $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and compared them with the exact parameter values from the GLD most closely approximating the exact distribution. There was emphasis given to comparing the tails of the distributions, since this is the most critical region in which the two distributions will differ. Even though the approximations are based on fourth moments, the estimated parameters for the GLD were shown to be quite close to the corresponding parameters of the original (exact) distribution.

For both symmetric and asymmetric distributions, the authors outlined the steps necessary to obtain estimates of $\lambda_1, \lambda_2, \lambda_3$ and λ_4 from a random sample. To assist

Table 7.1: Characteristics of Given Distributions. Reproduced from Ramberg and Schmeisser (1972,1974)

Distribution	λ_1	λ_2	λ_3	λ_4
Symmetric				
Normal (N)	0.0	.1975	.1349	.1349
t on 5 df (t_5)	0.0	-.3202	-.1359	-.1359
t on 10 df (t_{10})	0.0	.0261	.0148	.0148
t on 34 df (t_{34})	0.0	.1563	.1016	.1016
Uniform (U)	0.0	1.0000	.5774	.5774
Logistic (Lo)	0.0	-6.59E-4	-3.63E-4	-3.63E-4
Laplace (La)	0.0	-.1686	-.0802	-.0802
Cauchy (C)	0.0	-3.0674	-1.0000	-1.0000
Asymmetric				
Gamma Distribution: $F(x) = \beta^\alpha x^{\alpha-1}\epsilon^{-\beta x}/\Gamma(\alpha)$ $x > 0, \alpha > 1, \beta > 0$)				
$\alpha = 1, \beta = .5$ (G11)	.0008	.0002	.0000	.0004
$\alpha = 1, \beta = 1.0$ (G12)	.0004	.0002	.0000	.0004
$\alpha = 5, \beta = .5$ (G21)	7.1747	3.5874	.0252	.0939
$\alpha = 5, \beta = 1.0$ (G22)	.0216	.0434	.0252	.0939
$\alpha = 10, \beta = .5$ (G31)	16.7094	.0207	.0442	.1243
$\alpha = 10, \beta = 1.0$ (G32)	8.3547	.1243	.0442	.1243
$\alpha = 50, \beta = .5$ (G41)	95.8320	.0126	.0869	.1525
$\alpha = 50, \beta = 1.0$ (G42)	47.9160	.0251	.0869	.1525
$\alpha = 200, \beta = .5$ (G51)	395.4852	.0068	.1103	.1501
$\alpha = 200, \beta = 1.0$ (G52)	197.7426	.0136	.1103	.1501

in the computations, appropriate tables were produced.

For symmetric distributions, Ramberg and Schmeisser (1972) detailed the following steps to calculate estimates of the required λ combinations, given a random sample X_i, $i = 1, .., n$.

1. Compute

$$\hat{\theta} = \overline{X} = \sum_{i=1}^{n} X_i/n$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n,$$

$$\hat{\alpha}_4 = \sum_{i=1}^{n}(X_i - \overline{X})^4/n\hat{\sigma}^4$$

   from the data.

2. Find λ_2 and λ_3 satisfying the expressions for the central moments (or corresponding to the specified α_4 value in Ramberg and Schmeisser's (1972) Table I).

7.3.3 Ranking and selection using the GLD

Gupta and Sohn (1985) also used the GLD to examine the performance of ranking and selection procedures for selecting the population corresponding to the largest median. They, too, found this distribution gives satisfactory approximations to the exact distribution and showed that, with the lambda combination corresponding to the normal distribution (see Table 7.1), values of d obtained with the GLD satisfying $P_{k,1}(d) = 0.90, 0.95, 0.99$ agreed to at least two decimal places with the value of d obtained using the exact normal distribution.

Because of these satisfactory results, the problem to be considered in this chapter may be approached with the aid of the GLD. Using the value of d satisfying (7.2) for the normal distribution (tabulated in Table B.3) and assuming a common variance of unity, we wish to find the probability (7.3) for a non-normal distribution $F(x-\theta)$ which has been expressed in the form of a GLD. That is, we wish to investigate

how much variation can be tolerated in the lambda values of the GLD before the selection procedure P fails to give a satisfactory PCS.

As indicated in the introduction, this question is investigated by considering the lower bound P_L on the PCS given by (7.3). Notice that if P_L is satisfactorily close to the prespecified probability (P^*, say), the PCS may also be asserted to be close to (or larger than) P^*. If P_L is much larger than P^*, the PCS may be asserted to be even larger than P^*. If however, P_L is much smaller than P^*, it cannot be determined whether the bound is conservative (and the PCS is still satisfactory) or the PCS is also small. This problem is also discussed in a later section.

7.4 Computation of the lower bound for given distributions

To evaluate the expression $P_{k,t}(d)$ given by (7.4) for a given $(P^*, k, t, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ combination, the values of X satisfying (7.4) for p values of

$$\{.000(.001).005, .01(.01).99, .995(.001)1.00\}$$

were obtained. For each element of p, the value of $F(x + d)$ was then obtained by linear interpolation in the table of calculated values of X. The required probability $\hat{P}_L$ was approximated by replacing $P_{k,t}(d)$ in the right hand side of (7.3) by

$$\hat{P}_{k,t}(d) = t \sum (1 - F(x))^{t-1} (F(x + d))^{k-t} \Delta F(x)$$

(where Δ indicates forward differences), with summation over the values of p.

Using the GLD approximation to the normal distribution (given by (7.6), the $\hat{P}_{k,t}(d)$ values were compared with the corresponding exact $P_{k,t}(d)$ values tabulated by Bechhofer (1954) (also given in Table B.1 for $\nu = \infty$). In no case did the two figures differ by more than 0.001. This agrees with Gupta and Sohn's (1985) conclusions.

Using the value of d corresponding to $P^* = 0.75, 0.95$ for the normal distribution case (tabulated in Table B.3), the GLD representing other common distributions

(with parameters given in Table 7.1) was substituted for $F(\cdot)$ in (7.3) and the resulting value of $\hat{P}_L$ was obtained. This was done for a number of (k,t) combinations.

For most asymmetric distributions, the values of $\hat{P}_L$ were very low. (The only exception to this was G22, for which the values of $\hat{P}_L$ were quite close to the specified probability for all (P^*,k,t) combinations.) Many of the values were below the minimum possible PCS (using random selection of populations), indicating that it is the bound P_L that is extremely conservative in these cases. It is not possible from this investigation to infer whether the PCS is also low for these distributions (and hence whether Procedure P is robust to asymmetry.)

For most symmetric distributions, the $\hat{P}_L$ values were quite close to the specified probability P^*, for all (k,t) combinations. The exception to this was the Cauchy distribution. The $\hat{P}_L$ values obtained for symmetric distributions are given in Table 7.2, with the names of the distributions defined in Table 7.1.

Discussion of Table 7.2 For some reason, the combination $(k,t) = (2,1)$ appears to give results in Table 7.2 which are inconsistent with the general trends observed over the other (k,t) combinations. Although it is difficult to ascertain the reason for this, two factors may be main contributors. The first arises from the various approximations used in computing the tabulated d values (see Chapter 5), which are only accurate to ± 0.01 and in using the GLD instead of the exact distribution. The second consideration is the small size of the d value for this (k,t) combination; a change of 0.01 in this value may result in a fluctuation in $\hat{P}_L$ which is larger than that observed for the same change in a larger value of d. This case is not included in the following discussion on the trends observed in Table 7.2.

For the t distribution, the lower bound P_L (and hence the PCS) appeared to be quite close to P^*. Generally, there was much more variation in the $\hat{P}_L$ values at $P^* = 0.75$ compared with $P^* = 0.95$. For $P^* = 0.75$, all (k,t) combinations considered achieved $\hat{P}_L$ values which, when rounded to two decimal places, were greater than or equal to 0.75. (The smallest probability was 0.748.) For $P^* = 0.95$, more of the $\hat{P}_L$ were smaller than the specified P^*, with the smallest value being 0.92. Overall, for both P^* values, the procedure seemed to achieve the largest $\hat{P}_L$

Table 7.2: Values of $\hat{P}_L$ for selected symmetric distributions

(k,t)	P^*	t_5	t_{10}	$t_3 4$	U	Lo	La	C
(2,1)	.75	.72	.76	.75	.74	.76	.77	.66
(4,1)		.79	.76	.75	.72	.77	.78	.52
(7,1)		.78	.76	.75	.73	.76	.77	.40
(10,1)		.76	.75	.75	.75	.76	.76	.30
(4,2)		.80	.77	.75	.71	.77	.79	.42
(7,2)		.79	.76	.75	.71	.77	.78	.19
(10,2)		.77	.75	.75	.74	.76	.77	.00
(7,5)		.79	.76	.75	.72	.76	.77	.19
(10,5)		.78	.75	.75	.75	.76	.77	.00
(10,8)		.77	.75	.75	.75	.75	.76	.06
(2,1)	.95	.95	.95	.95	.95	.95	.95	.79
(4,1)		.95	.95	.95	.97	.95	.95	.65
(7,1)		.93	.94	.95	.99	.94	.94	.52
(10,1)		.93	.94	.95	1.00	.94	.93	.42
(4,2)		.94	.94	.95	.96	.95	.94	.59
(7,2)		.93	.94	.95	.99	.94	.93	.37
(10,2)		.92	.93	.95	1.00	.94	.93	.19
(7,5)		.93	.94	.95	.98	.94	.93	.37
(10,5)		.92	.93	.94	1.00	.93	.92	.14
(10,8)		.92	.93	.94	.99	.93	.92	.23

values for small k. These results tend to indicate that Procedure P is quite robust to heavier tails and in most cases gives a larger PCS (since this is always larger than P_L) than that required.

For the uniform distribution, the performance of the bound P_L appears to depend on the value of P^*. For $P^* = 0.75$, the values of $\hat{P}_L$ tended to be smaller than P^* (except for $k = 10$), whereas for $P^* = 0.95$, they were generally larger than P^* (except for $k = 2$). For larger P^*, then, it appears that Procedure P gives a much larger PCS than that required. For smaller P^* (such as 0.75), it can only be noticed that the PCS is at least greater than $\hat{P}_L$.

The logistic and Laplace distributions exhibited a quite different performance from that of the uniform, with most conservative results achieved for small k and t values. For $P^* = 0.75$, all observed $\hat{P}_L$ values were greater than 0.75, but for $P^* = 0.95$, all $\hat{P}_L$ values (with the exception of $k = 2$) were less than 0.95. The PCS obtained under these distributions, then, are likely to be greater than the required probability for $P^* = 0.75$ and not more than 0.03 below (and possibly greater than) the required probability for $P^* = 0.95$.

From Table 7.2, the Cauchy distribution performed very badly. Many of the $\hat{P}_L$ values were well below the minimum possible probability (if random selection of populations took place), indicating that the bound is extremely conservative in this situation. This investigation gives little information, then, about the true value of the PCS obtained for the Cauchy distribution.

7.4.1 Alternatives if P_L is small

As noticed above, if the P_L values are small, little may be asserted about the true PCS using the above approach. In such cases, two alternatives may be considered. Firstly, alternative bounds, such as $P_{k,t}(d)$ proposed by Carroll, Gupta and Huang (1975) or $P_{k-t+1,1}(d)$ conjectured by Bofinger and Mengersen (1986) and proved by Bofinger (1988) for $t = k - 1$, may be considered. It may be that these give larger bounds than $\hat{P}_L$ for the particular distribution. Certainly, they should form lower and upper bounds, respectively, for the PCS. A more appealing approach may be to simulate the true PCS, using the GLD to generate random variables from the

particular distribution, as described in Section 7.2.2.

To illustrate the simulation approach, Monte Carlo experiments were carried out for the following cases: uniform distribution, $P^* = 0.75$, $(k,t) = (4,1),(7,5)$; Cauchy distribution, $P^* = 0.95$, $(k,t) = (4,2),(7,5)$. Using an equal means configuration, the GLD approximating the particular distribution was used to simulate k random variables and the event of correct selection under Procedure P, using the appropriate value of d from Table B.3, was determined. This was repeated 500 times for $P^* = 0.75$ and 100 times for $P^* = 0.95$, giving observed values of the probability of correct selection ($\hat{PCS}$, say) with standard errors of 0.02. The resulting values of $\hat{PCS}$ for the four cases considered are given below.

Distribution	P^*	(k,t)	$\hat{P}_L$	$\hat{PCS}$
Uniform	.75	(4,22)	.71	.88
		(7,5)	.72	.73
Cauchy	.95	(4,2)	.72	.87
		(7,5)	.37	.68

7.5 General results

The results of Section 7.4 are particularly useful if the experimenter believes that the given data are not normal but that they do follow one of the distributions considered in that section. In this case, it is possible to determine if the use of the d values tabulated for the normal distribution will indeed give a satisfactory PCS-requirement even with the non- normal data.

Although this method of assessing robustness to non-normality by examining particular distributions is common in the literature, it has a disadvantage of not giving any information about the performance of the procedure when the data do not follow one of the examined distributions. Even if the data are "close" to one of the distributions studied, no real statement regarding the expected confidence level can be made.

Through use of the GLD, a more general assessment of robustness to non-normality may be made. By varying the parameters of the GLD, the $\hat{P}_L$ values may be calculated for the new λ combination, drawing a "picture" of the performance of the lower bound over the space of λ values. If only two λ parameters are allowed to vary, the results may be represented as a two-dimensional contour plot of probabilities or by three-dimensional perspective plots.

For symmetric distributions ($\lambda_3 = \lambda_4$), the behaviour of the bound as (λ_2, λ_3) vary was investigated in this way. For various (k, t) combinations, the value of d corresponding to a specified P^* value (given by 7.2 with $F(\cdot)$ the normal distribution function) was found from Table B.3. This (P^*, k, t, d) combination was then used in the calculation of $\hat{P}_L$ given by (7.3) with $F(\cdot)$ the GLD approximation to the particular distribution function under consideration). This was computed for each combination of $(\lambda_2, \lambda_3), \lambda_2 = -1.0(.05)1.0, \lambda_3 = -1.0(.05)1.0$, such that the sign of λ_2 is the same as the sign of λ_3 (since it is only in these two quadrants that the GLD is a valid density function). A selection of the resulting contour plots are displayed in Figure C.1 (located in Appendix C). These plots were generated using the statistical package S.

Discussion of Figure C.1 The sets of surfaces depicted in Figure C.1 include $(P^*, k, t) = (0.75,6,2), (0.95,6,2), (0.75,4,2), (0.95,4,2), (0.75,10,2), (0.75,10,5)$. The following overall observations may be made:

- For fixed k and t, as P^* increases the "cliffs" become more steep, and the two regions of zero probability (the surface "floor") and the valley connecting them (passing through the point (0,0)) become smaller. Consequently, the "ceiling" of the surface, representing $\hat{P}_L$ values close to unity, decreases in size.

- As k increases for fixed P^* and t, the surface "floor" area increases.

The observed performance of particular (symmetric) distributions, illustrated in Table 7.1, may be verified in these plots, by noting the "height" of the surface (the $\hat{P}$ value) for the appropriate (λ_2, λ_3) combination. Notice also that for the (λ_2, λ_3)

combination corresponding to the normal distribution, each of the plots gives a $\hat{P}_L$ value (almost) equal to P^*, as expected.

7.5.1 Empirical assessment of the effects of non-normality

For symmetric distributions, Figure C.1 may be used to assess the effect of non-normality on Procedure P for a particular data set. For asymmetric distributions, a series of similar plots may be generated, but since three parameters $(\lambda_2, \lambda_3, \lambda_4)$ must be considered, it may be easier to compute the approximate probability directly by using the estimated λ values in the GLD and substituting this distribution in (7.4).

A proposed method of using Figure C.1 is outlined below.

1. Test for symmetry in the data, by such methods as a stem-and-leaf display or normal probability plot. If there is an indication of asymmetry, consider possible transformations or stop and use an alternative method. If there is no asymmetry apparent, continue.

2. Estimate the values (λ_2, λ_3) using the methods suggested by Ramberg and Schmeisser (1972), described in Section 7.2.

3. Consider the plot in Figure C.1 corresponding to the particular (P^*, k, t) combination. Determine from the plot the value of $\hat{P}_L$ pertaining to the particular (λ_2, λ_3) combination.

   If this probability is close to P^* and not too conservative, use Procedure P with the d values corresponding to the normal distribution case, since the PCS will be close to (or greater than) P^* in this particular case (so that Procedure P is asserted to be robust to any non-normality present in the data).

   If the probability P_L is too small, a number of options are available. Firstly, consult the plots for the same (k, t) combination with smaller P^* value (P_2, say). Consider the $\hat{P}$ value corresponding to (k, t) from this plot. If this is satisfactorily close to P^*, use Procedure P with the value of d corresponding to $P^* = P_2$. This will result in a selection procedure which has PCS at least

P^*. Secondly, if the $\hat{P}_L$ value is still too small even for very large P_2, consider the options discussed earlier. These include evaluating alternative bounds to obtain more information about the size of the PCS or simulating the true PCS. Thirdly, it may be more desirable to consider another approach. For example, procedures designed for specific non-normal distributions (such as binomial) or nonparametric methods may be examined.

If the $\hat{P}_L$ values are too large, use the first option described for conservative values, but consult larger P^* values. Alternatively, consider other approaches, as discussed above.

The above methodology has the advantage of not requiring that the particular distribution be one of a small number of commonly investigated distributions. Even a slight deviation from normality may be assessed for its impact on the performance of the selection procedure P. Also, there is no requirement for tables of d values for distributions other than the normal; the d values for the normal means case may be applied for almost any continuous distribution. This inconvenience is replaced by another, though, of requiring contour plots for various (P^*, k, t) combinations. It may be possible to reduce the number of such plots required by investigating relationships between $\hat{P}_L$ values as P^*, k and t change (as was touched on in Section 7.3) and "interpolating" between plots. This is currently under consideration. If this is found to be feasible, it may be more useful to generate contour plots of the PCS values (rather than of the lower bound), obtained from simulation results. This would be practical only if it is possible to effectively reduce the number of contour plots required, since the computing time involved in simulating the PCS for each (λ_2, λ_3) combination for each (P^*, k, t) would be much greater than that required to obtain lower bounds on the PCS.

Notice also that this method may be extended to other selection procedures. The use of the GLD in the way described above enables quite general examination of the robustness of a wide range of procedures and the corresponding probability bounds.

7.6 Example

Consider the problem posed in Chapter 3, in which two lecturers are to be chosen as "best" from eight (two of whom were earlier excluded by other means). The selection is to be based on the scores of 21 students. From Chapter 3, the six remaining lecturers' final scores are:

Carole	Debbie	Wayne	Wendy	Russell	Janice
134.7	125.2	118.7	111.5	89.5	70.2

with $s^2 = 107.04$ on 140 degrees of freedom.

Suppose that the students wish to select, with at least 75% confidence, the two best lecturers. From Table B.3, the value of d corresponding to $(P^*, \nu, k, t) = (0.75,140,6,2)$ is 2.16. Hence, those lecturers with scores larger than $X_{R(5)} - ds/\sqrt{n} = 120.3$ are included in the selected subset; that is, Carole and Debbie are selected.

The above results were obtained under the assumption that the students' scores are independent and normally distributed with common (unknown) variance. Using the methods described in this chapter, we wish to investigate the validity of these assumptions and what effect any deviation from normality has on the confidence of correct selection. Although only the six remaining candidates will be considered in the final selection, the entire set of eight candidates' scores may be used in assessing the above assumptions.

Step 1 Using the residuals obtained after taking out lecturer and student effects from the set of scores $X_{uv}, u = 1,..,6, v = 1,..,21$ given in Chapter 3, we may test for homogeneity of variance, independence and symmetry in the distribution of the data.

In a two-way analysis of variance, the judge (in this case, student) effect is not significant ($F < 1$). Due to the large degrees of freedom for the estimated variance, we may assume that σ^2 is effectively known, but we must still ascertain homogeneity of variance. The estimated variances s_i^2 of the residuals of scores awarded to each lecturer $C_i, i = 1,..,8$ are computed to be:

David	Carole	Debbie	Wayne	Wendy	Chris	Russell	Janice
109.4	110.6	81.8	109.1	98.4	75.7	92.4	71.9.

The estimated variance s^2 computed from the combined set of residuals (from all eight lecturers) is 89.7. The chi-square statistic for a test of homogeneity of variance is 0.34410, which is non-significant. The range of correlations between lecturers is (-0.660,0.295), with 23 correlations negative and 5 positive.

Stem-and-leaf plots of the residuals for each lecturer, and the combined set of residuals, do not indicate asymmetry. Similarly, a plot of normal scores for the combined set of residuals is satisfactorily linear. For the combined data, a sign test for a median of zero gives a non-significant P-value. Estimates of skewness and kurtosis for the combined data are 0.48 and -1.49 respectively. The range of estimates of skewness for the individual lecturers is (-0.40,0.74) and of kurtosis is (-1.28,0.35).

Overall, then, the assumptions of homogeneity of variance and symmetry may be assumed to hold, but the assumption of independence may be in doubt. There is a suggestion of negative correlation between lecturers, but assume for the present that the data are independent so that we can proceed.

Step 2 The following may now be computed using the residuals:

$$\hat{\theta}(X) = \sum_{i=1}^{168} X_i = 0.0$$

$$\hat{\sigma}^2(X) = \sum_{i=1}^{168} (X_i - \overline{X})^2/168 = 89.2$$

$$\hat{\alpha}_4(X) = \sum_{i=1}^{168} (X_i - \overline{X})^4/(N\hat{\sigma}^4) = 2.47 \ .$$

In Procedure P, the means of 21 observations are used. Hence we need to consider the distribution of these means, rather than that of the individual observations. The moments of the standardised variables

$$Z = \overline{X}/\hat{\sigma}$$

are computed as follows:

$$\hat{\theta}(Z) = \hat{\theta}(X) = 0.0$$

$$\hat{\sigma}^2(Z) = \hat{\sigma}^2(X)/21 = 4.25$$

$$\hat{\alpha}_4(Z) = (\hat{\alpha}_4(X) + 3(21-1))/21 = 2.97 \ .$$

Interpolating in Ramberg and Schmeisser's Table I, we find that $(\lambda_2, \lambda_3) \doteq (.20, .14)$. (The values of (λ_2, λ_3) for the normal distribution are (.198,.135).)

Step 3 Consulting the contour plots in Figure C.1 at the point (λ_2, λ_3) for $(P^*, k, t) = (.75, 6, 2)$, we find that, if the percentage point corresponding to the latter combination in Table B.3 is used in Procedure P above, the lower bound $\hat{P}_L$ will be (as expected) very close to 0.75.

The students may then choose Carole and Debbie as the two best lecturers, with close to (if not greater than) 75% confidence of correct selection. They may be satisfied that this confidence holds even if their data are not exactly normally distributed. (The simulated value of the PCS in this case, using the methods described in Section 7.4.1), was found to be 0.85).

Notice that in the above approach, and for ranking and selection procedures in general, the tests of symmetry and homogeneity of variance applied above may not be the most suitable for this particular problem. Perhaps traditional tests of these assumptions are not entirely applicable for ranking and selection problems, since only a number of the populations are of interest. Alternatively, we could apply Ramberg and Schmeisser's (1974) steps for asymmetric distributions and compute the values of λ_3 and λ_4 separately. Some test of equality may then be applied to assess symmetry, which would help to determine the most appropriate approach. The development of suitable tests for symmetry and homogeneity for ranking and selection procedures requires further investigation.

It was also noticed above that there was some indication of negative correlation in the data. The largest correlations, however, were found to be between David and other lecturers. Remembering that David was excluded before any selection took place, it is difficult to assess the impact of such correlation. This question of the effect of correlation on ranking and selection procedures, and appropriate tests for correlation which may have adverse results on the procedures, requires further investigation and is at present under consideration.

Lastly, the means used in the above selection procedure were, as expected, much

more normally distributed than the individual observations. From the central limit theorem, it is reasonable to expect that, regardless of the underlying distribution, if a large enough sample is taken the means will be normally distributed. The size of such a sample is another problem requiring further consideration. It would also be interesting to assess the impact of this on the performance of asymmetric distributions.

Chapter 8

CONCLUSION

As stated in the introduction, this thesis attempts to provide answers to a number of selection problems. The goals considered in this thesis are those of comparison and selection of the t best populations, and simultaneous comparison with a control and the best. Although these occupy but a very small niche in the now quite large field of ranking and selection, they are quite fundamental problems. The methods developed to satisfy the goals employ the confidence bound approach, which provides an appealing alternative to the more traditional indifference zone and subset selection formulations.

Although the procedures developed to satisfy these ranking and selection goals have been demonstrated to satisfy the specified probability requirements and to be in some ways superior to existing methods, it is important not to ignore further questions that may be asked about the results. A real concern, for example, is that of the robustness of the procedures to deviation from any of the stated assumptions. One approach to such a question for the problem of selection of the t best means, if the assumption of normality is relaxed, was considered in Chapter 7 of this thesis. This approach may be extended to the methods developed in the other chapters to gain information about their robustness to normality. It remains, however, to investigate more fully the effect of relaxing any of the stated assumptions of the various selection procedures, including common variances and equal sample sizes and combinations of these, before the true practical story of these procedures'

robustness is fully understood.

It is the practical use of these methods that must be borne in mind, since ranking and selection, like all statistics, is primarily a tool to be used by practitioners. The theory of ranking and selection, while presently undergoing rapid development, has long ago matured to a level at which it is potentially relevant and applicable in practice. Although this is recognised by many researchers in the ranking and selection field, and although practitioners have been requesting effective solutions to the very questions that ranking and selection theory addresses, the combination of theory and application has not eventuated on nearly the scale expected.

Many reasons may be forwarded for this, although none of them are new. Lack of education of applied statisticians in the field of ranking and selection, both at university level and "on the job", has retarded the acceptance of the "new" methodology. This is exacerbated by the reluctance of researchers to publish relevant ranking and selection developments in journals commonly consulted by the applied practitioner, rather than in journals devoted solely to statistics. Often, however, even if practitioners are eager to embrace the methodology, the esoteric nature of the literature on the particular problem, the lack of relevant tables or, more commonly, the lack of computer software discourages the potential users.

This does not imply that no attempt has been made to apply ranking and selection theory in practice. In a large number of papers, real life or contrived problems are used to illustrate the application of the particular procedure. Some of these solutions are then presented in applied science journals. As discussed in Chapter 2, several reviews of ranking and selection have appeared in applied science journals and useful examples have been distributed throughout several books devoted to ranking and selection. Only recently, however, has the practical application been taken a step further, to embrace an entire practical problem in such a way that a user may follow each step of the methodology from the posing of a relevant question to the interpretation of the corresponding results.

In this thesis, such an attempt has been described. In the field of personnel selection, such goals as comparison, selection and ranking of candidates, goals that form the backbone of ranking and selection literature, are pertinent. In Chapter 3, a

ranking and selection methodology designed to be directly applicable to ranking and selection theory was detailed. Since it was not adequate to develop only the ranking and selection tools without first devising a quantitative scoring rule on which to base comparison between candidates, the methodology begins with the development of such a rule. Since it was also not adequate to expect all practitioners to manually execute the methodology, a computer package, PERSEL, was developed. PERSEL incorporates the goals and procedures of Chapters 4 and 5 of the thesis, as well as a number of other relevant ranking and selection goals.

As indicated, there are many potential improvements to be made and outstanding concerns about the selection and comparison procedures developed in this thesis. Bounds on the probability statements may be improved; questions such as robustness remain to be further investigated and application of the results to other related problems may also be made. So, too, there are many criticisms that may be made of the proposed application to personnel selection, including the scoring methodology, the advocated goals and procedures, and the package PERSEL. More sophisticated scoring methods may be introduced; a wider variety of goals and alternative procedures may be considered, and such improvements as graphical enhancement may be useful in PERSEL. All of these are currently under investigation.

The development of the theory of ranking and selection is undoubtedly most important, since it is on this solid base that practical application of the results may be made. The latter, however, is equally as important. It is from the continued interaction of researcher and practitioner that the field of ranking and selection will expand, with respect to the type of problems posed for solution, the scientific fields to which the theory is applied and the variety of people making useful contributions to the literature. In this light, it is not as important that the personnel selection methodology proposed in this thesis is not "best" in all respects, as the fact that it has been developed to directly answer specific questions in a specific field. From such a humble beginning, it is hoped that improvements and extensions will eventuate from both users and researchers, resulting in a statistically sophisticated, useful package.

It is also hoped that the field of personnel selection is not the only practical

area to benefit from the theoretical and practical developments presented in this thesis, but that the methodology may be extended to other related fields, such as market research, quality control and biometrics. With continued interaction between researchers and practitioners this may, in the not too distant future, be realised.