A Subset Selection Procedure for Selecting the Largest Multiple Correlation Coefficient

by

James N. Arvesen*

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series # 269

October 1971

* Research supported by the Purdue Research Foundation at Purdue University.

Abstract

A Subset Selection Procedure for Selecting the Largest Multiple Correlation Coefficient

Let $X' = (X_0, X_1 X_2)'$ have a multivariate normal distribution with arbitrary mean, covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{00} & \Sigma_{01} & \Sigma_{02} \\ \Sigma_{10} & \Sigma_{11} & \Sigma_{12} \\ \Sigma_{20} & \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ where } \Sigma_{00} \text{ is } 1 \times 1,$$

Σ_{11} is $t \times t$, Σ_{22} is $(p-t) \times (p-t)$. Let ρ^2_{max} denote the maximum of all $\binom{p}{t}$ possible multiple correlation coefficients. The goal is to obtain a subset selection procedure which contains ρ^2_{max}. The approach is Bayesian, using priors suggested by Geisser (Ann. Math. Statist. 36 150-159), and Lindley (Jour. Roy. Statist. Soc. 30 31-66). Let nS have a Wishart distribution with n degrees of freedom, and expectation $n\Sigma$. Then the asymptotic (as $n \to \infty$) joint posterior distribution of all $\binom{p}{t}$ multiple correlation coefficients is obtained. A result of Siotani (Kansas State University Technical Report No. 16, June, 1970) is used. The procedure reduces to a subset selection problem for means from a multivariate normal distribution.

1. Introduction. A recurring problem that a statistical consultant runs into is trying to obtain the "best" set of t variables out of p variables to predict another variable, say X_0. An excellent discussion of various techniques for the practitioner is found in Draper and Smith [1968]. Recently, in several papers, authors have recommended calculation of all sample multiple correlation coefficients, or at least enough to guarantee obtaining the largest sample multiple correlation coefficient (LaMotte and Hocking [1970], Furnival [1971]). One problem is to instruct the computer what to do with all the calculated multiple correlation coefficients. The present paper has the goal of obtaining a subset (of random size) of multiple correlation coefficients which contains the largest multiple correlation coefficient associated with a predictor based on t variables. The motivation is from Gupta's subset selection rules (see e.g. Gupta [1965]), but the approach is Bayesian. When costs of the predictor variables are known, an alternative Bayesian approach is given in Lindley [1968]. The present paper treats only the asymptotic case.

2. Some distribution theory. Siotani [1971] considers the problem of obtaining the asymptotic joint distribution of $\binom{p}{t}$ sample multiple correlation coefficients between a variate and t variates taken from p possible variates, $t < p$. More specifically, $Z = (X_0, X_1, \ldots, X_p)$ are assumed to have a $(p+1)$-variate normal distribution with covariance matrix

$$\sum = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdot & \cdot & \cdot & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \cdot & \cdot & \cdot & \sigma_{1p} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \sigma_{p0} & \sigma_{p1} & \cdot & \cdot & \cdot & \sigma_{pp} \end{bmatrix} .$$

$S = (s_{ij})$, $i, j = 0, 1, \ldots, p$ is the sample covariance matrix, nS being Wishart with covariance matrix $\sum$ and n degrees of freedom. Let $i = (i_1, \ldots, i_t)$ $j = (j_1, \ldots, j_t)$,

$$\sum_{(i)}^{0} = (\sigma_{0i_1}, \ldots, \sigma_{0i_t}), \qquad \sum_{0}^{(i)} = (\sum_{(i)}^{0})' ,$$

$$\sum_{(i)} = \begin{bmatrix} \sigma_{i_1 i_1} & \cdot & \cdot & \cdot & \sigma_{i_1 i_t} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \sigma_{i_t i_1} & & & & \sigma_{i_t i_t} \end{bmatrix} ,$$

$$\sum_{0(i)} = \begin{bmatrix} \sigma_{00} & \sum_{(i)}^{0} \\ \sum_{0}^{(i)} & \sum_{(i)} \end{bmatrix} ,$$

$$\Sigma_{(j)}^{(i)} = \begin{bmatrix} \sigma_{i_1 j_1} & \cdot & \cdot & \cdot & \sigma_{i_1 j_t} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \sigma_{i_t j_1} & \cdot & \cdot & \cdot & \sigma_{i_t j_t} \end{bmatrix}$$

and corresponding notation for S. Then

$$(2.1) \qquad \rho_{0(i)}^2 = \frac{\Sigma_{(i)}^0 \, \Sigma_{(i)}^{-1} \, \Sigma_0^{(i)}}{\sigma_{00}} = 1 - \frac{|\Sigma_{0(i)}|}{\sigma_{00} \, |\Sigma_{(i)}|} ,$$

with corresponding notation for $r_{0(i)}^2$, the sample multiple correlation coefficient.

Then Siotani shows that the joint distribution of the $\binom{p}{t}$ $r_{0(i)}^2$'s is asymptotically normal as $n \to \infty$ with means $\rho_{0(i)}^2$, and

$$(2.2) \qquad \mathrm{var}(r_{0(i)}^2) \cong \frac{4}{n} \rho_{0(i)}^2 \, (1-\rho_{0(i)}^2)^2 .$$

The covariance between two squared coefficients $r_{0(i)}^2$ and $r_{0(j)}^2$ will now be given.

Suppose $(i_1,\ldots,i_t)$ and $(j_1,\ldots,j_t)$ have $r(< t)$ indices in common, and we let $(i,k) = (i_1,\ldots,i_{t-r}, k_1,\ldots,k_r)$, $(k,j) = (k_1,\ldots,k_r, j_1,\ldots,j_{t-r})$, and from now on $(i) = (i_1,\ldots,i_{t-r})$, $(j) = (j_1,\ldots,j_{t-r})$, $(k) = (k_1,\ldots,k_r)$. Then, following Siotani,

$$\mathrm{Cov}(r^2_{0(i,k)}, r^2_{0(k,j)}) \cong \frac{2}{n}(1-\rho^2_{0(i,k)})(1-\rho^2_{0(k,j)})$$

$$\begin{aligned}(2.3)\quad [\rho^2_{0(i,k)}+\rho^2_{0(k,j)} &- \Sigma^{-1}_{0.(k,j)}\Sigma^{0}_{(i).(k,j)}\Sigma^{-1}_{(i).(k)}\Sigma^{(i)}_{0.(k)} \\ &- \Sigma^{-1}_{0.(i,k)}\Sigma^{0}_{(j).(i,k)}\Sigma^{-1}_{(j).(k)}\Sigma^{(j)}_{0.(k)} \\ &+ \Sigma^{-1}_{0.(k,j)}\Sigma^{0}_{(i).(k,j)}\Sigma^{-1}_{(i).(k)}\Sigma^{(i)}_{0.(k)}\Sigma^{-1}_{0.(i,k)}\Sigma^{0}_{(j).(i,k)}\Sigma^{-1}_{(j).(k)}\Sigma^{(j)}_{0.(k)}]\ .\end{aligned}$$

$$\begin{aligned}&= \frac{2}{n}(1-\rho^2_{0(i,k)})(1-\rho^2_{0(k,j)})\ [\rho^2_{0(i,k)} + \rho^2_{0(k,j)} \\ &- (1-\rho^2_{0(k,j)})^{-1}\,\frac{\Sigma^{0}_{(i).(k,j)}\Sigma^{-1}_{(i).(k)}\Sigma^{(i)}_{0.(k)}}{\sigma_{00}} \\ (2.4)\quad &-(1-\rho^2_{0(i,k)})^{-1}\,\frac{\Sigma^{0}_{(j).(i,k)}\Sigma^{-1}_{(j).(k)}\Sigma^{(j)}_{0.(k)}}{\sigma_{00}} + (1-\rho^2_{0(k,j)})^{-1}\,(1-\rho^2_{0(i,j)})^{-1}\ . \\ &\cdot\frac{\Sigma^{0}_{(i).(k,j)}\Sigma^{-1}_{(i).(k)}\Sigma^{(i)}_{0.(k)}}{\sigma_{00}}\cdot\frac{\Sigma^{0}_{(j).(i,k)}\Sigma^{-1}_{(j).(k)}\Sigma^{(j)}_{0.(k)}}{\sigma_{00}}\,]\ .\end{aligned}$$

Note that

$\Sigma^{0}_{(i).(k,j)} = \Sigma^{0}_{(i)} - \Sigma^{0}_{(k,j)}\Sigma^{-1}_{(k,j)}\Sigma^{(k,j)}_{(i)}$ is the covariance between 0 and (i) in the conditional distribution given (k,j), and similarly for the other conditional expressions. Siotani also obtains another representation for (2.4) which will not be used here.

Next, the asymptotic joint distribution of partial multiple correlation coefficients will be obtained. Let

$$\text{(2.5)}\quad \rho^2_{0(i).(i)^c} = \frac{\Sigma^{0}_{(i).(i)^c}\,\Sigma^{-1}_{(i).(i)^c}\,\Sigma^{(i)}_{0.(i)^c}}{\sigma_{00.(i)^c}} = 1 - \frac{|\Sigma_{0(i).(i)^c}|}{\sigma_{00.(i)^c}\,|\Sigma_{(i).(i)^c}|}$$

where $\Sigma_{0(i).(i)^c} = \Sigma_{0(i)} - \Sigma^{0(i)}_{(i)^c}\,\Sigma^{-1}_{(i)^c}\,\Sigma^{(i)^c}_{0(i)}$, and similarly for the other terms, and $(i)^c = (1,\ldots,p) \sim (i_1,\ldots,i_t)$. Following Siotani, the joint asymptotic distribution of the $\binom{p}{t}$ $r^2_{0(i).(i)^c}$'s is normal with means $\rho^2_{0(i).(i)^c}$, and

$$\mathrm{Var}(r^2_{0(i).(i)^c}) \cong \frac{4}{n}\,\rho^2_{0(i).(i)^c}\,(1-\rho^2_{0(i).(i)^c})^2$$

As before, the bulk of the work is to obtain the covariance between $r^2_{0(i).(i)^c}$ and $r^2_{0(j).(j)^c}$. Let (i,k) and (j,k) be as described above, and let (g) represent the $p - 2t + r$ indices left. In what follows, an approach similar to Siotani's will be used, and it will be assumed that the reader is familiar with this approach.

$$\begin{aligned}
&\mathrm{Cov}(r^2_{0(i,k).(i,k)^c},\ r^2_{0(k,j).(k,j)^c}) \\
&= \mathrm{Cov}(r^2_{0(i,k).(j,g)},\ r^2_{0(k,j).(i,g)}) \\
&\cong (1-\rho^2_{0(i,k).(j,g)})(1-\rho^2_{0(k,j).(i,g)})\left(\frac{\mathrm{Cov}(s_{00.(j,g)},\ s_{00.(i,g)})}{\sigma_{00.(j,g)}\ \sigma_{00.(i,g)}}\right. \\
\text{(2.6)}\quad &+ \frac{\mathrm{Cov}(s_{00.(j,g)},\ |S_{(k,j).(i,g)}|)}{\sigma_{00.(j,g)}\ |\Sigma_{(k,j).(i,g)}|} - \frac{\mathrm{Cov}(s_{00.(j,g)},\ |S_{0(k,j).(i,g)}|)}{\sigma_{00.(j,g)}\ |\Sigma_{0(k,j).(i,g)}|} \\
&+ \frac{\mathrm{Cov}(|S_{(i,k).(j,g)}|,\ s_{00.(i,g)})}{|\Sigma_{(i,k).(j,g)}|\,\sigma_{00.(i,g)}} + \frac{\mathrm{Cov}(|S_{(i,k).(j,g)}|,\ |S_{(k,j).(i,g)}|)}{|\Sigma_{(i,k).(j,g)}|\ |\Sigma_{(k,j).(i,g)}|}
\end{aligned}$$

$$- \frac{\mathrm{Cov}(|S_{(i,k).(j,g)}|, |S_{0(k,j).(i,g)}|)}{|\Sigma_{(i,k).(j,g)}| \; |\Sigma_{0(k,j).(i,g)}|} - \frac{\mathrm{Cov}(|S_{0(i,k).(j,g)}|, s_{00.(i,g)})}{|\Sigma_{0(i,k).(j,g)}| \; \sigma_{00.(i,g)}}$$

$$- \frac{\mathrm{Cov}(|S_{0(i,k).(j,g)}|, |S_{(k,j).(i,g)}|)}{|\Sigma_{0(i,k).(j,g)}| \; |\Sigma_{(k,j).(i,g)}|} + \frac{\mathrm{Cov}(|S_{0(i,k).(j,g)}|, |S_{0(k,j).(i,g)}|)}{|\Sigma_{0(i,k).(j,g)}| \; |\Sigma_{0(k,j).(i,g)}|}$$

In an attempt to simplify some of the notation in what follows, let us now define some quantities to be used in subsequent calculations. Let

$$\alpha_{ji} = \Sigma^{-1}_{0.(j,g)} \Sigma^{0}_{(i).(j,g)} ,$$

$$\alpha_{jk} = \Sigma^{-1}_{0.(j,g)} \Sigma^{0}_{(k).(j,g)} ,$$

$$\alpha_{ij} = \Sigma^{-1}_{0.(i,g)} \Sigma^{0}_{(j).(i,g)}$$

$$\beta_{ji} = - \Sigma^{-1}_{0.(j,g)} \Sigma^{-1}_{(j).(g)} \Sigma^{(j)}_{0.(g)} \Sigma^{0}_{(i).(j,g)} + \Sigma^{-1}_{(j).(g)} \Sigma^{(j)}_{(i).(g)} ,$$

$$\beta_{ij} = - \Sigma^{-1}_{0.(i,g)} \Sigma^{-1}_{(i).(g)} \Sigma^{(i)}_{0.(g)} \Sigma^{0}_{(j).(i,g)} + \Sigma^{-1}_{(i).(g)} \Sigma^{(i)}_{(j).(g)} ,$$

$$\delta_{ji} = \Sigma^{-1}_{(j).(g)} \Sigma^{(j)}_{(i).(g)} , \tag{2.7}$$

$$\delta_{ij} = \Sigma^{-1}_{(i).(g)} \Sigma^{(i)}_{(j).(g)} ,$$

$$\gamma_{j} = \Sigma^{-1}_{(j).(g)} \Sigma^{(j)}_{0.(g)}$$

$$\gamma_{i} = \Sigma^{-1}_{(i).(g)} \Sigma^{(i)}_{0.(g)} ,$$

$$\varepsilon_{i} = \Sigma^{-1}_{(i).(j,g,k)} \Sigma^{(i)}_{0.(j,g,k)} ,$$

$$\varepsilon_{k} = \Sigma^{-1}_{(k).(j,g,i)} \Sigma^{(k)}_{0.(j,g,i)} .$$

$$\varepsilon_{j} = \Sigma^{-1}_{(j).(i,g,k)} \Sigma^{(j)}_{0.(i,g,k)} ,$$

$$\alpha_{ik} = \Sigma^{-1}_{0.(i,g)} \Sigma^{0}_{(k).(i,g)} .$$

For what follows, the covariance in (2.6) will be computed from the full covariance matrix Σ, which will be written in the form,

$$\Sigma = \begin{bmatrix} \sigma_{00} & \Sigma^{0}_{(j)} & \Sigma^{0}_{(g)} & \Sigma^{0}_{(i)} & \Sigma^{0}_{(k)} \\ \Sigma^{(j)}_{0} & \Sigma_{(j)} & \Sigma^{(j)}_{(g)} & \Sigma^{(j)}_{(i)} & \Sigma^{(j)}_{(k)} \\ \Sigma^{(g)}_{0} & \Sigma^{(g)}_{(j)} & \Sigma_{(g)} & \Sigma^{(g)}_{(i)} & \Sigma^{(g)}_{(k)} \\ \Sigma^{(i)}_{0} & \Sigma^{(i)}_{(j)} & \Sigma^{(i)}_{(g)} & \Sigma_{(i)} & \Sigma^{(i)}_{(k)} \\ \Sigma^{(k)}_{0} & \Sigma^{(k)}_{(j)} & \Sigma^{(k)}_{(g)} & \Sigma^{(k)}_{(i)} & \Sigma_{(k)} \end{bmatrix}$$

$$= P \begin{bmatrix} \sigma_{00} & \Sigma^{0}_{(i)} & \Sigma^{0}_{(g)} & \Sigma^{0}_{(k)} & \Sigma^{0}_{(j)} \\ \Sigma^{(i)}_{0} & \Sigma_{(i)} & \Sigma^{(i)}_{(g)} & \Sigma^{(i)}_{(k)} & \Sigma^{(i)}_{(j)} \\ \Sigma^{(g)}_{0} & \Sigma^{(g)}_{(i)} & \Sigma_{(g)} & \Sigma^{(g)}_{(k)} & \Sigma^{(g)}_{(j)} \\ \Sigma^{(k)}_{0} & \Sigma^{(k)}_{(i)} & \Sigma^{(k)}_{(g)} & \Sigma_{(k)} & \Sigma^{(k)}_{(j)} \\ \Sigma^{(j)}_{0} & \Sigma^{(j)}_{(i)} & \Sigma^{(j)}_{(g)} & \Sigma^{(j)}_{(k)} & \Sigma_{(j)} \end{bmatrix} P^{-1}$$

for some permutation matrix P.

Now, the various covariances in (2.6) will be evaluated using the technique described in Siotani. First

$\mathrm{Cov}\,(s_{00.(j,g)}, s_{00.(i,g)})$

$$= \mathrm{Cov}\left(\frac{\begin{vmatrix} s_{00} & S^{0}_{(j,g)} \\ S^{(j,g)}_{0} & S_{(j,g)} \end{vmatrix}}{|S_{(j,g)}|} , \frac{\begin{vmatrix} s_{00} & S^{0}_{(i,g)} \\ S^{(i,g)}_{0} & S_{(i,g)} \end{vmatrix}}{|S_{(i,g)}|} \right)$$

$$(2.8) \qquad \cong \frac{2}{n} \sigma_{00.(j,g)} \sigma_{00.(i,g)} \mathrm{tr}(\Phi_1 \textstyle\sum \Psi_1 \sum)$$

where

$$\Phi_1 = \begin{bmatrix} \sum^{-1}_{0(j,g)} & 0 \\ & \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum^{-1}_{(j,g)} & 0 \\ 0 & 0 & 0 \end{bmatrix} ,$$

$$\Psi_1 = P \left\{ \begin{bmatrix} \sum^{-1}_{0(i,g)} & 0 \\ & \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum^{-1}_{(i,g)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} P^{-1} .$$

Thus $\Phi_1 \sum \Psi_1 \sum =$

$$\left\{ \begin{bmatrix} I_{p-t+1} & \sum^{-1}_{0(j,g)} \sum^{0(j,g)}_{(i,k)} \\ & \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ \sum^{-1}_{(j,g)} \sum^{(j,g)}_{0} & I_{p-t} & \sum^{-1}_{(j,g)} \sum^{(j,g)}_{(i,k)} \\ 0 & 0 & 0 \end{bmatrix} \right\}$$

$$.P\left\{\begin{bmatrix} I_{p-t+1} & \Sigma^{-1}_{0(i,g)} & \Sigma^{0(i,g)}_{(k,j)} \\ & & \\ 0 & & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ \Sigma^{-1}_{(i,g)}\Sigma^{(i,g)}_{0} & I_{p-t} & \Sigma^{-1}_{(i,g)}\Sigma^{(i,g)}_{(k,j)} \\ 0 & 0 & 0 \end{bmatrix}\right\} P^{-1}$$

$$= \left\{\begin{bmatrix} 1 & 0 & 0 & \alpha_{ji} & * \\ 0 & I_{t-r} & 0 & \beta_{ji} & * \\ 0 & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_j & I_{t-r} & 0 & \delta_{ji} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\}$$

$$P\left\{\begin{bmatrix} 1 & 0 & 0 & * & \alpha_{ij} \\ 0 & I_{t-r} & 0 & * & \beta_{ij} \\ 0 & 0 & I_{p+2t-r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_j & I_{t-r} & 0 & * & \delta_{ij} \\ * & 0 & I_{p+2t-r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\} P^{-1}$$

Thus

$$\mathrm{tr}\ (\Phi_1 \Sigma \Psi_1 \Sigma) =$$

$$(2.9) \quad 1 + \mathrm{tr}\ (\beta_{ji}\ \beta_{ij} - \alpha_{ji}\ \gamma_i - \beta_{ji}\ \delta_{ij} - \gamma_j\ \alpha_{ij} - \delta_{ji}\ \beta_{ij} + \delta_{ji}\ \delta_{ij})$$

Next, we obtain

$$\mathrm{Cov}\left(\frac{s_{00.(j,g)}}{\sigma_{00.(j,g)}} \;,\; \frac{|S_{(k,j).(i,g)}|}{|\Sigma_{(k,j).(i,g)}|}\right)$$

$$(2.10) \qquad \cong \frac{2}{n}\ \mathrm{tr}(\Phi_1 \Sigma \Psi_2 \Sigma) \;,$$

where

$$\Psi_2 = P\left\{\begin{bmatrix} 0 & 0 \\ & \\ 0 & \Sigma^{-1}_{(i,g,k,j)} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma^{-1}_{(i,g)} & 0 \\ 0 & 0 & 0 \end{bmatrix}\right\} P^{-1} \;.$$

Thus $(\Phi_1 \Sigma \Psi_2 \Sigma) =$

$$\left\{\begin{bmatrix} 1 & 0 & 0 & \alpha_{ji} & \alpha_{jk} \\ 0 & I_{t-r} & 0 & \beta_{ji} & * \\ 0 & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_j & I_{t-r} & 0 & \delta_{ji} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\}$$

$$P\left\{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \varepsilon_i & I_{t-r} & 0 & 0 & 0 \\ * & 0 & I_{p-2t+r} & 0 & 0 \\ \varepsilon_k & 0 & 0 & I & 0 \\ * & 0 & 0 & 0 & I \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_j & I_{t-r} & 0 & * & \delta_{ij} \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\} P^{-1}$$

and hence

$$\mathrm{tr}\ (\Phi_1 \textstyle\sum \Psi_2 \sum) =$$

(2.11) $$\mathrm{tr}(\alpha_{ji}\ \varepsilon_i + \alpha_{jk}\ \varepsilon_k - \alpha_{ji}\ \gamma_i - \beta_{ji}\ \delta_{ij} + \delta_{ji}\ \delta_{ij})$$

Next, we obtain

$$\mathrm{Cov}\left(\frac{s_{00.(j,g)}}{\sigma_{00.(j,g)}}\ ,\ \frac{|S_{0(k,j).(i,g)}|}{|\Sigma_{0(k,j).(i,g)}|}\right)$$

(2.12) $$\simeq \frac{2}{n}\ \mathrm{tr}(\Phi_1 \textstyle\sum \Psi_3 \sum)\ ,$$

where

$$\Psi_3 = P\left\{\Sigma^{-1} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma^{-1}_{(i,g)} & 0 \\ 0 & 0 & 0 \end{bmatrix}\right\} P^{-1}\ .$$

Thus

$$(\Phi_1 \textstyle\sum \Psi_3 \sum)$$

$$= \left\{\begin{bmatrix} I & 0 & 0 & \alpha_{ji} & * \\ 0 & I_{t-r} & 0 & \beta_{ji} & * \\ 0 & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_i & I_{t-r} & 0 & \delta_{ji} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\}$$

$$\cdot\ p\left\{\begin{bmatrix} & & \\ & I_{p+1} & \\ & & \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \gamma_i & I_{t-r} & 0 & * & \delta_{ij} \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\}$$

and hence

$$\mathrm{tr}\ (\Phi_1 \textstyle\sum \Psi_3 \sum)$$

$$(2.13)\quad = 1 + \mathrm{tr}(-\ \alpha_{ji}\ \gamma_i - \beta_{ji}\ \delta_{ij} + \delta_{ji}\ \delta_{ij}).$$

Now

$\mathrm{Cov}\left(\frac{|S_{(i,k).(j,g)}|}{|\sum_{(i,k).(j,g)}|}\ ,\ \frac{s_{00.(i,g)}}{\sigma_{00.(i,g)}}\right)$ can be obtained from (2.10) and (2.11) by interchanging (i) and (j), thus

$$\mathrm{Cov}\left(\frac{|S_{(i,k).(j,g)}|}{|\sum_{(i,k).(j,g)}|}\ ,\ \frac{s_{00.(i,g)}}{\sigma_{00.(i,g)}}\right)$$

$$(2.14)\quad \simeq \frac{2}{n}\ \mathrm{tr}(\alpha_{ij}\ \epsilon_j + \alpha_{ik}\ \epsilon_k - \alpha_{ij}\ \gamma_j - \beta_{ij}\ \delta_{ji} + \delta_{ij}\ \delta_{ji})\ .$$

Next, we obtain

$$\mathrm{Cov}\left(\frac{|S_{(i,k).(j,g)}|}{|\sum_{(i,k).(j,g)}|}\ ,\ \frac{|S_{(k,j).(i,g)}|}{|\sum_{(k,j).(i,g)}|}\right)$$

(2.15) $\simeq \frac{2}{n} \; tr(\Phi_2 \sum \Psi_2 \sum)$ where

$$\Phi_2 = \left\{ \begin{bmatrix} 0 & 0 \\ & \\ 0 & \sum^{-1}_{(j,g,i,k)} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum^{-1}_{(j,g)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} .$$

Thus

$\Phi_2 \sum \Psi_2 \sum =$

$$\left\{ \begin{bmatrix} 0 & 0 \\ & \\ * & I_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & \delta_{ji} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\} .$$

$$\cdot P \left\{ \begin{bmatrix} 0 & 0 \\ & \\ * & I_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & * & \delta_{ij} \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\} P^{-1} ,$$

and

(2.16) $tr\;(\Phi_2 \sum \Psi_2 \sum) = (t-r) + \delta_{ji}\,\delta_{ij}$.

Next, we obtain

$$\operatorname{Cov}\left(\frac{|S_{(i,k).(j,g)}|}{|\Sigma_{(i,k).(j,g)}|}, \frac{|S_{0(k,j).(i,g)}|}{|\Sigma_{0(k,j).(i,g)}|}\right)$$

(2.17) $\simeq \frac{2}{n}$ tr $(\Phi_2 \Sigma \Psi_3 \Sigma)$, where

$$\Phi_2 \Sigma \Psi_3 \Sigma = \left\{\begin{bmatrix} 0 & 0 \\ & \\ 0 & I_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & \delta_{ji} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\}$$

$$P\left\{\begin{bmatrix} I_{p+1} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & * & \delta_{ij} \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right\} P^{-1},$$

and thus

(2.18) $\operatorname{tr}(\Phi_2 \Sigma \Psi_3 \Sigma) = (t-r) + \operatorname{tr}(\delta_{ji}\, \delta_{ij})$

Now

$$\mathrm{Cov}\left(\frac{|S_{0(i,k).(j,g)}|}{|\Sigma_{0(i,k).(j,g)}|}\ ,\ \frac{s_{00.(i,g)}}{\sigma_{00.(i,g)}}\right)$$

$$(2.19)\quad \cong \frac{2}{n}\ (1 + tr(-\alpha_{ij}\ \gamma_j - \beta_{ij}\ \delta_{ji} + \delta_{ij}\ \delta_{ji}))$$

as obtained from (2.13) by interchanging i and j. Also

$$\mathrm{Cov}\left(\frac{|S_{0(i,k).(j,g)}|}{|\Sigma_{0(i,k).(j,g)}|}\ ,\ \frac{|S_{(k,j).(i,g)}|}{|\Sigma_{(k,j).(i,g)}|}\right)$$

$$(2.20)\quad \cong \frac{2}{n}\ \{\ (t-r) + tr(\delta_{ij}\ \delta_{ji})\ \}$$

as seen by interchanging i and j in (2.17) and (2.18).

Finally

$$\mathrm{Cov}\left(\frac{|S_{0(i,k).(j,g)}|}{|\Sigma_{0(i,k).(j,g)}|}\ ,\ \frac{|S_{0(k,j).(i,g)}|}{|\Sigma_{0(k,j).(i,g)}|}\right)$$

$$(2.21)\quad \cong \frac{2}{n}\ tr(\phi_3\ \Sigma\ \psi_3\ \Sigma)\quad \text{where}$$

$$\Phi_3 \sum \Psi_3 \sum = \left\{ \begin{bmatrix} I_{p+1} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & \delta_{ij} & * \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\}$$

$$P^{-1} \left\{ \begin{bmatrix} I_{p+1} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ * & I_{t-r} & 0 & * & \delta_{ji} \\ * & 0 & I_{p-2t+r} & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\} P \quad ,$$

and thus

(2.22) $$\mathrm{tr}(\Phi_3 \textstyle\sum \Psi_3 \sum) \cong (1 + (t-r) + \delta_{ij}\, \delta_{ji}) \quad .$$

Combining the appropriate terms into (2.6) one obtains,

(2.23)
$$\begin{aligned} &\mathrm{Cov}\,(r^2_{0(i,k).(j,g)},\ r^2_{0(k,j).(i,g)}) \\ &\cong \frac{2}{n}\,(1-\rho^2_{0(i,k).(j,g)})(1-\rho^2_{0(k,j).(i,g)})[- \textstyle\sum^{-1}_{0.(j,g)} \sum^{0}_{(i).(j,g)} \sum^{-1}_{(i).(g)} \sum^{(i)}_{0.(g)} \\ &- \textstyle\sum^{-1}_{0.(i,g)} \sum^{0}_{(j).(i,g)} \sum^{-1}_{(j).(g)} \sum^{(j)}_{0.(g)} + \sum^{-1}_{0.(j,g)} \sum^{0}_{(i).(j,g)} \sum^{-1}_{(i).(g)} \sum^{(i)}_{0.(g)} \\ &\textstyle\sum^{-1}_{0.(i,g)} \sum^{0}_{(j).(i,g)} \sum^{-1}_{(j).(g)} \sum^{(j)}_{0.(g)} \\ &+ \textstyle\sum^{-1}_{0.(j,g)} \sum^{0}_{(i,k).(j,g)} \sum^{-1}_{(i,k).(j,g)} \sum^{(i,k)}_{0.(j,g)} \\ &+ \textstyle\sum^{-1}_{0.(i,g)} \sum^{0}_{(j,k).(i,g)} \sum^{-1}_{(j,k).(i,g)} \sum^{(j,k)}_{0.(i,g)}\ ] \quad . \end{aligned}$$

Note that the last two terms in the bracket above are $\rho^2_{0(i,k).(j,g)}$ and $\rho^2_{0(k,j).(i,g)}$ respectively. Thus (2.23) may be written as

$$\mathrm{Cov}\,(r^2_{0(i,k).(j,g)},\, r^2_{0(j,g).(i,k)})$$

$$\cong \frac{2}{n}\,(1-\rho^2_{0(i,k).(j,g)})(1-\rho^2_{0(k,j).(i,g)})[\rho^2_{0(i,k).(j,g)} + \rho^2_{0(k,j).(i,g)}$$

$$\text{(2.24)}\quad - (1-\rho^2_{0(j,g)})^{-1}\,\frac{\Sigma^{0}_{(i).(j,g)}\Sigma^{-1}_{(i).(g)}\Sigma^{(i)}_{0.(g)}}{\sigma_{00}} - \frac{(1-\rho^2_{0(i,g)})^{-1}\Sigma^{0}_{(j).(i,g)}\Sigma^{-1}_{(j).(g)}\Sigma^{(j)}_{0.(g)}}{\sigma_{00}}$$

$$+ (1-\rho^2_{0(j,g)})^{-1}(1-\rho^2_{0(i,g)})^{-1}\,\frac{\Sigma^{0}_{(i).(j,g)}\Sigma^{-1}_{(i).(g)}\Sigma^{(i)}_{0.(g)}}{\sigma_{00}}\,\frac{\Sigma^{0}_{(j).(i,g)}\Sigma^{-1}_{(j).(g)}\Sigma^{(j)}_{0.(g)}}{\sigma_{00}}$$

For subsequent work, we will leave the covariance in this form.

3. Selection of the largest multiple correlation coefficient.

a. Background. Let us define the best set of t predictor variables to be that set $X_{i_1},\ldots,X_{i_t}$, $(i) = (i_1,\ldots,i_t)$ such that

$$\text{(3.1)}\qquad \rho^2_{\max} = \max_{(j)} \rho^2_{0(j)}$$

where the max is taken over all $\binom{p}{t}$ sets. Note that the set $X_{i_1},\ldots,X_{i_t}$ may not be unique, in which case we will consider them as an equivalence class and define any one of them best. One procedure to select the set (i) (and hence the associated $X_{i_1},\ldots,X_{i_t}$) is the use of the subset selection approach advocated in a general setting by Gupta [1965].

This latter approach has a serious practical drawback in the present case of being too conservative. To illustrate this, let $p = 2$ and let the goal be to select a subset that contains the predictor with associated ρ^2_{max} with probability at least $(1-\alpha)$. Note that the "practical" value of such a rule would be to make one of the following three statements:

(3.2)
(i.) X_1 is the best predictor or
(ii.) X_2 is the best predictor or
(iii.) it is impossible to state whether X_1 is better than X_2 and conversely.

These three statements are subject to the probability requirement under all possible parameter configurations.

The type of rule suggested by Gupta's general approach is of the form

R: include in the selected subset those predictors X_i, $i = 1, 2$ with

$$r^2_{0(i)} \geq c \max_{(j)} r^2_{0(j)} \quad . \tag{3.3}$$

To obtain c, $0 < c < 1$, one must solve

$$P\left(\frac{S_{0(i)} S^{-1}_{(i)} S_{(i)0}}{S_{00}} \geq c \max_{(j)} \frac{S_{0(j)} S^{-1}_{(j)} S_{(j)0}}{S_{00}}\right) \geq 1-\alpha \quad \text{when}$$

$\rho^2_{(1)(2)} = \rho^2_{0(1)} = \rho^2_{0(2)} = 0$ and hence solve $P(F_{1,1} \geq c) \geq 1-\alpha$ where $F_{1,1}$ is an F statistic with 1 and 1 degrees of freedom. For example, if $\alpha = .10$, $c = (39.9)^{-1}$, and thus in almost every practical situation, statement (iii) of (3.2) would be made.

Nevertheless, the subset selection procedure as given by Gupta has been successfully applied in numerous other problems ([1956], [1958], [1962]). It fails here because of the underlined words after (3.2). To overcome this difficulty, a natural thing to do is to make probability statements conditional on the actual data observed, not on some hypothetical parameter configuration that has little practical relevance. An excellent argument for this point of view is found in Hoadley [1970].

To achieve this goal, let $\sum^{-1}$ have the prior distribution given by

(3.4) $$p(\textstyle\sum^{-1}) \propto |\textstyle\sum|^{(p+2)/2} \quad ,$$

and obtain the result that

(3.5) $$\textstyle\sum^{-1} \mid S^{-1} \sim \mathcal{W}(n, S^{-1}) \quad ,$$

that is $n \sum^{-1}$ has a Wishart distribution with covariance matrix S^{-1} (positive definite with probability one) and n degrees of freedom. The prior given in (3.4) has been considered and discussed adequately in Geisser and Cornfield [1963], Geisser [1965], Lindley [1965], and Enis and Geisser [1971]. Finally, we note that the actual form of the prior given in (3.4) should make little difference in the instant problem since asymptotic results are being obtained as $n \to \infty$.

Let

$$\Lambda = \Sigma^{-1} \begin{bmatrix} \sigma^{00} & \cdot & \cdot & \cdot & \sigma^{0p} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \sigma^{p0} & \cdot & \cdot & \cdot & \sigma^{pp} \end{bmatrix}$$

$\Lambda_{(i)}$, $\Lambda^{0}_{(i)}$, $\Lambda_{0(i)}$, etc. defined analogous to corresponding terms for Σ in section 2.

Let

$$(3.6) \qquad (\rho^{0(i)})^2 = \frac{\Lambda^{0}_{(i)} \Lambda^{-1}_{(i)} \Lambda^{(i)}_{0}}{\sigma^{00}} ,$$

$$(\rho^{0(i).(i)^c})^2 = \frac{\Lambda^{0}_{(i).(i)^c} \Lambda^{-1}_{(i).(i)^c} \Lambda^{(i)}_{0.(i)^c}}{\sigma^{00.(i)^c}}$$

Then the following lemma is needed.

Lemma 1.

(i.) $\rho^2_{0(i).(i)^c} = (\rho^{0(i)})^2$,

(ii.) $\rho^2_{0(i)} = (\rho^{0(i).(i)^c})^2$.

PROOF. (i.) First note that

$$\rho^2_{0(i).(i)^c} = \frac{\Sigma^0_{(i).(i)^c}\, \Sigma^{-1}_{(i).(i)^c}\, \Sigma^{(i)}_{0.(i)^c}}{\sigma_{00.(i)^c}} .$$

But following Anderson [1958], pp. 341-344,

$$\Sigma^{-1}_{(i).(i)^c} = \Lambda_{(i).0} ,$$

$$\Sigma^0_{(i).(i)^c} = -(\sigma^{00})^{-1}\, \Lambda^0_{(i)}\, \Lambda^{-1}_{(i).0} ,$$

and $\sigma_{00.(i,i^c)} = (\sigma^{00})^{-1}$.

Thus

$$\frac{\rho^2_{0(i).(i)^c}}{1-\rho^2_{0(i).(i)^c}} = \frac{\Sigma^0_{(i).(i)^c}\, \Sigma^{-1}_{(i).(i)^c}\, \Sigma^{(i)}_{0.(i)^c}}{\sigma_{00.(i,i^c)}}$$

$$= \frac{\Lambda^0_{(i)}\, \Lambda^{-1}_{(i).0}\, \Lambda^{(i)}_0}{\sigma^{00}} .$$

But $\Lambda^{-1}_{(i).0} = (\Lambda_{(i)} - \Lambda^{(i)}_0\, (\sigma^{00})^{-1}\, \Lambda^0_{(i)})^{-1}$

$$= \Lambda^{-1}_{(i)} + [\sigma^{00}(1-\rho^{0(i)})^2]^{-1}\, \Lambda^{-1}_{(i)}\, \Lambda^{(i)}_0\, \Lambda^0_{(i)}\, \Lambda^{-1}_{(i)} ,$$

and thus

$$\frac{\rho^2_{0(i).(i)^c}}{1-\rho^2_{0(i).(i)^c}} = (\rho^{0(i)})^2 + (1-\rho^{0(i)^2})^{-1}\, (\rho^{0(i)})^4 = \frac{(\rho^{0(i)})^2}{1-(\rho^{0(i)})^2} .$$

The proof of (ii.) is similar.

Let $T = S^{-1} = \begin{bmatrix} s^{00} & . & . & . & s^{0p} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ s^{p0} & . & . & . & s^{pp} \end{bmatrix}$. We now state the following theorems.

Theorem 1. For the prior distribution given in (3.4), the joint asymptotic posterior distribution of the $\binom{p}{t}$ multiple regression coefficients $\rho^2_{0(i,k)}$ is multivariate normal with means $(r^{0(i,k).(j,g)})^2$ variances given by

(3.7) $$\mathrm{Var}(\rho^2_{0(i,k)}|\text{data}) \cong \frac{4}{n} (r^{0(i,k).(j,g)})^2 (1-(r^{0(i,k).(j,g)})^2)^2 ,$$

and

$$\mathrm{Cov}(\rho^2_{0(i,k)}, \rho^2_{0(k,j)}|\text{data}) \cong \frac{2}{n} (1-(r^{0(i,k).(j,g)})^2)(1-(r^{0(k,j).(i,g)})^2) .$$

(3.8) $$\cdot \Big[(r^{0(i,k).(j,g)})^2 + (r^{0(k,j).(i,g)})^2 - (1-(r^{0(j,g)})^2)^{-1} \frac{T^0_{(i).(j,g)} T^{-1}_{(i).(g)} T^{(i)}_{0.(g)}}{s^{00}}$$

$$- (1-(r^{0(i,g)})^2)^{-1} \frac{T^0_{(j).(i,g)} T^{-1}_{(j).(g)} T^{(j)}_{0.(g)}}{s^{00}}$$

$$+ (1-(r^{0(j,g)})^2)^{-1} (1-(r^{0(i,g)})^2)^{-1} \frac{T^0_{(i).(j,g)} T^{-1}_{(i).(g)} T^{(i)}_{0.(g)}}{s^{00}} .$$

$$\cdot \frac{T^0_{(j).(i,g)} T^{-1}_{(j).(g)} T^{(j)}_{0.(g)}}{s^{00}} \Big] .$$

Theorem 2. For the prior distribution given in (3.4), the joint asymptotic posterior distribution of the $\binom{p}{t}$ partial multiple regression coefficients $\rho^2_{0(I,k).(j,g)}$ is multivariate normal with means $(r^{0(i,k)})^2$, variances given by

(3.9) $$\mathrm{Var}(\rho^2_{0(i,k).(j,g)} \mid \text{data}) \simeq \frac{4}{n} (r^{0(i,k)})^2 (1-(r^{0(i,k)})^2)^2 , \text{ and}$$

$$\mathrm{Cov}(\rho^2_{0(i,k).(j,g)}, \rho^2_{0(k,j).(i,g)} \mid \text{data}) \simeq \frac{2}{n} (1-(r^{0(i,k)})^2)(1-(r^{0(k,j)})^2) \; .$$

$$\cdot \Big[(r^{0(i,k)})^2 + (r^{0(k,j)})^2 - (1-(r^{0(k,j)})^2)^{-1} \frac{T^0_{(i).(k,j)} \; T^{-1}_{(i).(k)} \; T^{(i)}_{0.(k)}}{s^{00}}$$

$$- (1-(r^{0(i,k)})^2)^{-1} \frac{T^0_{(j).(i,k)} \; T^{-1}_{(j).(k)} \; T^{(j)}_{0.(k)}}{s^{00}}$$

$$+ (1-(r^{0(k,j)})^2)^{-1} (1-(r^{0(i,j)})^2)^{-1} \frac{T^0_{(i).(k,j)} \; T^{-1}_{(i).(k)} \; T^{(i)}_{0.(k)}}{s^{00}} \; .$$

$$\cdot \frac{T^0_{(j).(i,k)} \; T^{-1}_{(i).(k)} \; T^{(j)}_{0.(k)}}{s^{00}} \Big] \; .$$

PROOF. The proofs of the two theorems are similar, and only theorem 1 will be given, and it has a straightforward proof. Note that

$\rho^2_{0(i,k)} = (\rho^{0(i,k).(j,g)})^2$ by lemma 1. But interchanging ρ and r, also $\sum$ and $T = S^{-1}$, the joint asymptotic posterior distribution of the

$\binom{p}{t}$ $(\rho^{0(i,k).(j,g)})^2$ statistics was obtained in section 2, the covariance being found in (2.24). Thus the result follows.

b. The proposed subset selection rule. The goal of this section is to obtain a rule for subset selection that yields a subset (of random size) of multiple correlation coefficients that contains ρ^2_{max} with probability at least $1-\alpha$, where the probability is calculated with respect to the asymptotic joint posterior distribution given in theorem 1. Hence the problem can be formulated in the following terms:

Consider the posterior distribution for μ, an $m \times 1$ vector,

$$\mu|\text{data} \sim \mathcal{N}_m(\tau, U) \tag{3.10}$$

where τ is an $m \times 1$ known mean vector, U is a positive semi-definite $m \times m$ known covariance matrix. Assume $\tau' = (\tau_1,\ldots,\tau_m)$ is such that $\tau_1 \geq \tau_2 \geq \ldots \geq \tau_m$. The goal is to select a subset (of random size), that contains with posterior probability at least $1-\alpha$

$$\bar{\mu} = \max_{1 \leq i \leq m} \mu_i \tag{3.11}$$

where $\mu' = (\mu_1,\ldots,\mu_m)$.

The proposed rule is of the form R: include $\mu_1,\ldots,\mu_\ell$ in the selected subset, where ℓ is the first index such that

$$P(\max(\mu_1,\ldots,\mu_\ell) \geq \max(\mu_{\ell+1},\ldots,\mu_m)|\text{data}) \geq 1-\alpha, \tag{3.12}$$

where the distribution given in theorem 1 is used in (3.10). In practice, one might prefer to calculate (3.12) for several values of ℓ.

Note that for the original problem, $m = \binom{p}{t}$, and thus (3.12) may prove difficult to evaluate on a computer, at least in closed form. One possible solution is to obtain (3.12) by Monte Carlo techniques. For example, let

$p = 6$, $t = 3$, and assume a standard error of at most .01 is allowable for (3.12). Then 2500 sets of $\binom{6}{3} = 20$, or 50,000 pseudo normal deviates are needed, a not insurmountable task. Larger problems would require simplification.

The author is constructing computer programs to implement the proposed selection rule. Finally, it should be again emphasized that the proposed procedure is discussed in the asymptotic case. For moderate samples, one may prefer a transformation of the parameters (e.g. $\tanh^{-1}(\rho_{0(i)})$).

REFERENCES

[1] Anderson, T.W. [1958]. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] Draper, N.R. and Smith, H. [1968]. Applied Regression Analysis. Wiley, New York.

[3] Enis, P. and Geisser, S. [1971]. Estimation of the probability that Y < X. Jour. Am. Stat. Assn. 66 162-168.

[4] Furnival, G. [1971]. All possible regressions with less computation. Technometrics 13 403-408.

[5] Geisser, S. [1965]. Bayesian estimation in multivariate analysis. Ann. Math. Statist. 36 150-159.

[6] Geisser, S., and Cornfield, J. [1963]. Posterior distributions for multivariate normal parameters. Jour. Roy. Statist. Soc. Ser. B 25 368-376.

[7] Gupta, S.S. [1956]. On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

[8] Gupta, S.S., and Sobel, M. [1958]. On selecting a subset which contains all populations better than a standard. Ann. Math. Statist. 29 235-244.

[9] Gupta, S.S., and Sobel, M. [1962]. On selecting a subset containing the population with the smallest variance. Biometrika 49 495-507.

[10] Gupta, S.S. [1965]. On some multiple decision (selection and ranking) rules. Technometrics 7 225-245.

[11] Hoadley, B. [1970]. A Bayesian look at inverse linear regression. Jour. Am. Stat. Assn. 65 356-369.

[12] La Motte, L., and Hocking, R. [1970]. Computational efficiency in the selection of regression variables. Technometrics 12 83-94.

[13] Lindley, D. [1968]. The choice of variables in multiple regression (with discussion). Jour. Roy. Statist. Soc. Ser. B 30 31-66.

[14] Siotani, M. [1971]. Asymptotic joint distribution of $\binom{p}{t}$ multiple correlation coefficients between a certain variate and t variates among p other variates (t < p). Technical Report No. 16, Kansas State University, Manhattan, Kansas.