Subset Selection Problems for Variances with Applications to Regression Analysis

By J. N. Arvesen and G. P. McCabe, Jr.*

Summary

This paper presents a subset selection procedure for correlated variances. Emphasis is placed on the asymptotic case. An application to selecting the best set of independent variables in a regression problem is given.

Some key words: Subset selection; regression analysis; correlated variances.

1. Introduction

In Gupta and Sobel [4], the following problem is considered. Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ denote k normal populations with unknown variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$. Let $\sigma_{[1]}^2 \leq \sigma_{[2]}^2 \leq \cdots \leq \sigma_{[k]}^2$ denote the ordered variances (equalities are allowed for mathematical convenience only). The goal is to select a subset of the k populations which contains the best population, where the best population is defined to be the one associated with $\sigma_{[1]}^2$. Any such selection is called a correct selection (CS).

*J. N. Arvesen is associate director for scientific affairs, Pfizer Pharmaceuticals, Inc., 235 East 42nd Street, New York, NY 10017 and adjunct assistant professor, Columbia University, NY. G. P. McCabe is assistant professor and head of statistical consulting, Department of Statistics, Purdue University, West Lafayette, IN 47907. This research was supported in part by Office of Naval Research Contract N00014-67-A-0226-00014 and National Science Foundation Grant GJ30269 at Purdue University.

For each population, an independent sample of size n is used to calculate a sample variance, s_i^2. Let $s_{[1]}^2 \leq s_{[2]}^2 \leq \cdots \leq s_{[k]}^2$ denote the ordered sample variances. Gupta and Sobel have proposed rules of the following form:

Retain Π_i in the selected subset if and only if

$$s_i^2 \leq s_{[1]}^2/c, \tag{1.1}$$

where $0 < c \leq 1$. Given n,k and P*, a value of c can be determined so that the probability of correct selection, P(CS), is at least P* for any possible configuration of the parameters $\sigma_1^2,\ldots,\sigma_k^2$. Tables of c are given in Gupta and Sobel [5].

In the present paper, a rule of the form given in (1.1) is proposed for the case where the sample variances are correlated. Section 2 defines the problem and necessary notation. Section 3 treats the special case k = 2, and a table of some comparisons with the Gupta and Sobel results is given. In Section 4, arbitrary values of k are considered and a method for obtaining the c values of (1.1) is proposed. Section 5 applies the results of the previous sections to the regression analysis problem of selecting the best subset of independent predictor variables for any given subset size.

2. The Selection Problem

Let $W' = (W_1', W_2', \ldots, W_k')$ be a normally distributed random vector of length kn with mean zero and covariance matrix Σ. Each W_i is a random vector of length n with

$$E(W_i W_i') = \sigma_i^2\, \Sigma_{ii} = \sigma_i^2\, I,$$

and

$$E(W_i W_j') = \sigma_i \sigma_j\, \Sigma_{ij},$$

with Σ_{ij} positive definite. Thus $\Sigma = (\sigma_i\sigma_j\ \Sigma_{ij})$, and may be singular. Denote the ordered variances by $\sigma^2_{[1]} \leq \ldots \leq \sigma^2_{[k]}$. The goal then is to find a rule R that satifies (1.1). As in Gupta and Sobel [4], we consider rules of the form retain Π_i in the selected subset if and only if

$$SS_i \leq SS_{[1]}/c, \qquad i=1,\ldots,k, \tag{2.1}$$

where $SS_i = W_i'\ W_i$, and $SS_{[1]} \leq SS_{[2]} \leq \ldots \leq SS_{[k]}$.

Let $SS_{(i)}$ denote the sum of squares associated with $\sigma^2_{[i]}$. Note that SS_i is associated with W_i or equivalently with population Π_i, whereas $SS_{[i]}$ is the i-th smallest sum of squares and $SS_{(i)}$ is the sum of squares corresponding to the (unknown) i-th smallest expected sum of squares $\sigma^2_{[i]}$. As before, let P(CS) denote the probability that the population corresponding to $\sigma^2_{[1]}$ is included in the selected subset. Thus,

$$\begin{aligned} P(CS) &= P(SS_{(1)} \leq SS_{[1]}/c) \\ &= P(SS_{(1)} \leq SS_{(j)}/c,\ j=2,\ldots,k) \\ &= P(c(\sigma^2_{[1]}/\sigma^2_{[j]})(SS_{(1)}/\sigma^2_{[1]}) \leq SS_{(j)}/\sigma^2_{[j]},\ j=2,\ldots,k) \\ &\geq P(c\ SS_{(1)}/\sigma^2_{[1]} \leq SS_{(j)}/\sigma^2_{[j]},\ j=2,\ldots,k), \end{aligned} \tag{2.2}$$

where the inequality follows from the fact that $\sigma^2_{[1]}/\sigma^2_{[j]} \leq 1$, for $j=2,\ldots,k$. Furthermore, it is clear that the bound in (2.2) approaches a minimum value as the parameters $\sigma^2_{[j]}$ approach $\sigma^2_{[1]}$. Since this limiting probability does not depend on the value of $\sigma^2_{[1]}$, we can and do assume $\sigma^2_{[1]} = 1$ in what follows.

Thus,

$$P(CS) = P(cSS_{(1)} \leq SS_{(j)},\ j=2,\ldots,k), \tag{2.3}$$

where $SS_i = V_i'V_i$, with $V' = (V_1', V_2', \ldots, V_k')$ normally distributed with mean vector zero and covariance matrix $\Sigma = (\Sigma_{ij})$, and $\Sigma_{ii} = I$. Note that the random variables SS_i are marginally chi-square with n degrees of freedom. Thus, the problem of calculating a lower bound P* for P(CS) involves the joint distribution of a set of dependent chi-square random variables. One should note that we are considering the case Σ_{ij} known, and for the application in Section 5 this is the case. Finally, one should note that the right hand side of (2.3) is not in general invariant under all permutations of subscripts. This problem is treated in Section 4.

It is straightforward (see Krishnamoorthy and Parthasarathy [8]) to show that the joint characteristic function of the $\{SS_j/2\}$ is given by

$$\begin{aligned} \varphi(t_1, \ldots, t_k) &= E[\exp(i \, \Sigma_{j=1}^{k} t_j \, SS_j/2)] \\ &= |I - i\Sigma \, T|^{-1/2} \qquad (2.4) \end{aligned}$$

where $T = (\mathrm{diag}(t_1, \ldots, t_k)) \otimes I_n$. In principle, the function above can be used to find the joint density function. Integration over the appropriate set would then yield the bound for P*. In practice, however, such a computation presents considerable analytic difficulty. Special cases, on the other hand, can be treated. In the next 2 sections some possible approaches are discussed.

3. The Case of Two Sums of Squares

This section serves a two-fold purpose. First, the problem for k = 2 has interest in itself (see, e.g. Hotelling [6]). Secondly, the mathematical difficulties encountered in this simple case are indicative of the greater difficulties present in the higher dimensional cases and thus represent a justification for using asymptotic methods and for suggesting Monte Carlo techniques in later sections.

When $k = 2$, the problem can be reduced to a relatively simple form by transforming to canonical variates. Thus, we can assume that $\Sigma_{11} = \Sigma_{22} = I$ and $\Sigma_{12} = \Sigma_{21} = \mathrm{diag}(\rho_1, \ldots, \rho_n)$. With this transformation, Jensen [7] has obtained the joint density of $u_1 = SS_1/2$ and $u_2 = SS_2/2$ as

$$f(u_1, u_2) = \Psi(u_1)\Psi(u_2)\Sigma_{m=0}^{\infty} h_m L_m^{(n/2-1)}(u_1) L_m^{(n/2-1)}(u_2) \tag{3.1}$$

where

$$h_m = G_m(\rho_1, \ldots, \rho_n)\ \{m!\ \Gamma(n/2)/\Gamma(n/2+m)\}^2,$$

and

$$G_m(\rho_1, \ldots, \rho_n) = \Sigma_{j_1+\ldots+j_n=m} \Pi_{i=1}^{n} \rho_i^{2j_i} \Gamma(j_i + \tfrac{1}{2})/\{\Gamma(j_i+1)\Gamma(\tfrac{1}{2})\},$$

where the outer sum is over all integer partitions of m, and

$$\Psi(u) = u^{(n/2-1)} e^{-u}/\Gamma(n/2).$$

The functions $L(u)$ are Laguerre polynomials,

$$L_m^{(g-1)}(u) = \Sigma_{i=0}^{m} (-1)^i u^i \Gamma(m+g)/\{\Gamma(m-i+1)\Gamma(g+i)\Gamma(i+1)\}.$$

Since the density is symmetric in u_1 and u_2, a solution can be obtained by setting

$$P^* = P(cSS_1 \leq SS_2)$$

$$= \Sigma_{m=0}^{\infty} h_m \int_0^{\infty}\int_{cu_1}^{\infty} L_m^{(n/2-1)}(u_1) L_m^{(n/2-1)}(u_2)\Psi(u_1)\Psi(u_2)du_1 du_2. \tag{3.2}$$

Since this expression is a strictly decreasing function of c, the solution can be obtained by iterative methods on a computer. Note that the terms to be integrated are polynomials which do not depend on the $\{\rho_i\}$. However, if some of the $\{\rho_i\}$ are large, the convergence in (3.2) will be slow.

In the special case $\rho_1 = \ldots = \rho_n = \rho$, some further simplification is possible. Gunst and Webster [3] present arguments which suggest that this configuration may provide a useful approximation for unequal ρ_i. Using the results of Siotani [13], it can be shown that

$$P(cSS_1 \leq SS_2) = \Sigma_{\ell=0}^{\infty} P(L=\ell) P\{F(n+2\ell, n+2\ell) \leq 1/c\} \tag{3.3}$$

where $F(n+2\ell, n+2\ell)$ denotes an F random variable with $n+2\ell$, and $n+2\ell$ degrees of freedom and L in a compound Poisson variable with parameters $1-\rho^2$ and $n/2$. Letting

$$F(n+2\ell, n+2\ell, c) = P\{F(n+2\ell, n+2\ell) \leq 1/c\},$$

one obtains the useful approximation inequalities

$$\Sigma_{\ell=0}^{m} P(L=\ell) F(n+2\ell, n+2\ell, c) + (1-\Sigma_{\ell=0}^{m} P(L=\ell)) F(n+2m, n+2m, c)$$

$$< \Sigma_{\ell=0}^{\infty} P(L=\ell) F(n+2\ell, n+2\ell, c)$$

$$< \Sigma_{\ell=0}^{m} P(L=\ell) F(n+2\ell, n+2\ell, c) + (1-\Sigma_{\ell=0}^{m} P(L=\ell)). \tag{3.4}$$

Alternatively, if n is large, an Edgeworth approximation can be derived.

Table I gives values calculated for $1/c$ with $P^* = .90$. Formula (3.3) was used for $n = 4$ and 10 while an Edgeworth approximation was used for $n = 30$ and 50. The $\rho^2 = 0$ column corresponds to the Gupta and Sobel [5] case. From this table the increased sensitivity gained by increasing the squared correlation is evident. While a table for various values of n and ρ is possible to construct, perhaps a computer program is preferable. In fact, tables would be essentially impossible in the general case when the $\{\rho_i\}$ are unequal. The situation becomes even more complicated for arbitrary k.

INSERT TABLE I ABOUT HERE

Table 1. Values of $1/c$ for selected combinations of n and ρ^2 ($P^* = .90$)

n \ ρ^2	0	.25	.50	.75	.90	1.00
4	4.11	3.47	2.83	2.14	1.63	1.00
10	2.32	2.08	1.84	1.54	1.32	1.00
30	1.61	1.51	1.40	1.27	1.16	1.00
50	1.44	1.37	1.30	1.20	1.12	1.00

4. Asymptotic Case for Arbitrary k

In Chambers [1], an Edgeworth type approximation is obtained for problems such as the one at hand. It is necessary to obtain the joint cumulant generating function of $SS_1/2,\ldots,SS_k/2$. Note that the cumulant generating function based on (2.4) is (following Searle [12]),

$$\log|I-i\Sigma T|^{-\frac{1}{2}} = \frac{1}{2}\Sigma_{r=1}^{\infty} i^r tr(\Sigma T)^r/r$$

$$= \frac{1}{2}\Sigma_{r=1}^{\infty} i^r C_r(t_1,\ldots,t_k)/r. \qquad (4.1)$$

Thus, the joint cumulant $K_{r_1 r_2,\ldots,r_k}$ of total order $r = r_1+r_2+\ldots+r_k$, can be obtained from the rth term of (4.1) by multiplying the coefficient of $i^r t_1^{r_1} t_2^{r_2} \ldots t_k^{r_k}$ by $r_1!\ldots r_k!$ Note that for $r = 1,2,3$,

$$C_1 = n(\Sigma_{j=1}^{k} t_j)/2,$$

$$C_2 = n\{\Sigma_{j=1}^{k} t_j^2 + 2\Sigma_{i<j} t_i t_j tr(\Sigma_{ij}\Sigma_{ji})\}/2,$$

and

$$C_3 = (2n/3)\{\Sigma_{j=1}^{k} t_j^3 + 3\Sigma_{i\neq j} t_i^2 t_j tr(\Sigma_{ij}\Sigma_{ji})$$

$$+ 6\Sigma_{h<i<j} t_h t_i t_j tr(\Sigma_{hi}\Sigma_{ij}\Sigma_{jh})\}. \qquad (4.2)$$

Carrying terms this far enables one to make an Edgeworth approximation to order $n^{-1/2}$.

Let us return to the original problem of (2.3), that is, given Σ, find that configuration of $SS_{(1)},\ldots,SS_{(k)}$ such that

$$P(cSS_{(1)} \leq SS_{(j)},\ j = 2,\ldots,k) \geq P^* \qquad (4.3)$$

when n is large. To this end, note that upon standardizing SS_i, (4.3) becomes

$$P(cZ_{(1)} \le Z_{(j)} + (n/2)^{1/2}(1-c),\ j=2,\ldots,k), \tag{4.4}$$

where

$$Z_{(i)} = (SS_{(i)}-n)/(2n)^{1/2},\ i=1,\ldots,k)$$

The covariance matrix of the $\{Z_{(i)}\}$ is given by $\Gamma=(\rho_{ij})$ where

$$\rho_{ij} = n^{-1}tr(\Sigma_{ij}\Sigma_{ji}),\ i \neq j$$

and

$$\rho_{ii} = 1.$$

Let $\tau_{(i)}$ denote the conditional variance of $Z_{(i)}$ given $Z_{(1)},\ldots,Z_{(i-1)},Z_{(i+1)},\ldots,Z_{(k)}$. Note that the $\{\tau_{(i)}\}$ are functions of the known matrix Γ. When n is large, $(n/2)^{1/2}(1-c)$ is large and we apply the normal approximation. Thus, conditioning on $Z_{(2)},\ldots,Z_{(k)}$, one sees that (4.4) is minimized when $\tau_{(1)}$ is the largest of the k conditional variances. Thus the criterion

$$P(CS) \geq P^* \tag{4.5}$$

is achieved for n sufficiently large by the following rule:

(i) Calculate all the k conditional variances $\{\tau_{(i)}\}$, $i=1,\ldots,k$, assuming by relabeling if necessary that $\tau_{(1)}$ is the largest.

(ii) Then solve for c,

$$P(cZ_{(1)} \le Z_{(j)} + (n/2)^{\frac{1}{2}}(1-c),\ j=2,\ldots,k) = P^*$$

using the Edgeworth approximation, where $Z_{(i)}=(SS_{(i)}-n)/(2n)^{1/2}$, $i=1,\ldots,k$.

(iii) Then retain Π_i in the selected subset if and only if

$$SS_i \le SS_{[1]}/c. \tag{4.6}$$

In practice for k greater than four or five, this may be a formidable problem. For k large, one may have to resort to Monte Carlo techniques, rather than rely on the Edgeworth approximation. One slight simplification is possible by considering the k-1 random variables $T_{(i)}=cZ_{(1)}-Z_{(i)}$, $i=2,\ldots,k$.

5. Selecting the Best Regression Equation of Size $t < p$.

Assume the following standard linear model,

$$Y = X\beta + \varepsilon, \tag{5.1}$$

where X is an Nxp known matrix of rank $p \leq N$, β is a px1 parameter vector, and $\varepsilon \sim N(0,\sigma^2 I_N)$.

In what follows, equation (5.1) which includes p independent variables, will be viewed as the "true" model. For various reasons, however, one may be interested in including only a subset (say of size $t < p$) of the independent variables. Various authors have considered this problem and a variety of techniques are presently being used to construct such subsets. (see e.q. [6], [14] and references in [10])

La Motte and Hocking [9] have developed an algorithm for obtaining optimal subsets in the sense of maximum sample multiple correlations. This algorithm does not require calculation of all the $\binom{p}{t}$ possibilities. Furnival [2] on the other hand, has developed a method for efficiently calculating all possible multiple correlations. These procedures appear to be practical for values of p up to about 25.

Determination of the subset of size t which maximizes the multiple correlation or equivalently minimizes the residual sum of squares is not equivalent to finding the subset which is optimal in terms of expected values of these quantities. In this section, a procedure for taking into account the statistical variation of the residual sum of squares is proposed.

Consider the modesl

$$Y = X_i\beta_i + \varepsilon_i, \tag{5.2}$$

where X_i is an Nxt matrix (of rank t), β_i is a tx1 parameter vector, and $\varepsilon_i \sim N(0,\sigma_i^2 I_N)$, where $i=1,\ldots,k=\binom{p}{t}$ over all possible partitions. The goal is to select that design X_i (or set of independent variables) associated with $\sigma^2_{[1]}$ where $\sigma^2_{[1]} \leq \ldots \leq \sigma^2_{[k]}$. We will consider rules of the Gupta and Sobel form as given in (2.1), and use the rule given in (4.6).

Note that if

$$SS_i = Y'\{I-X_i(X_i'X_i)^{-1}X_i'\}Y = Y'Q_iY, \tag{5.3}$$

then following Searle (1972, p. 57),

$$SS_i/\sigma^2 \sim \chi^2\,\{n,(X\beta)'Q_i(X\beta)/(2\sigma^2)\},$$

where n = N-t. Note that the non-centrality parameter is not zero in general and that $\sigma_i^2=\sigma^2 + (X\beta)'Q_i(X\beta)/n$. Again, since rules of the form (2.1) or (4.6) are invariant with respect to $\sigma^2 > 0$, we assume without loss of generality that $\sigma^2=1$.

To obtain the joint distribution of $SS_1,\ldots,SS_k$, we can write

$$Y'Q_iY = U_i'\,U_i,$$

where

$$U_i = B_iY,$$

and

$$B_i B_i' = I,$$

$$B_i' B_i = Q_i, \tag{5.4}$$

where B_i is an nxN matrix.

The joint distribution of $U' = (U_1', \ldots, U_k')$ is multivariate normal in kn dimensions, with mean vector $\eta' = (\eta_1', \ldots, \eta_k')$ with $\eta_i = B_i X\beta$, and covariance matrix $\Sigma = (\Sigma_{ij})$ where $\Sigma_{ij} = B_i B_j'$ is nxn. Note that the knxkn covariance matrix Σ is possibly singular. Let $\Sigma = FF'$ where F is of full column rank r(r=rank (Σ)), and let $U = \eta + FA$, where $A \sim N(0, I_r)$. Thus the joint characteristic function of $SS_1/2, \ldots, SS_k/2$ is (since $SS_i = U_i' U_i$),

$$\varphi(t_1, \ldots, t_k) = E\{\exp\,(i\Sigma_{j=1}^{k} t_j U_j' U_j/2)\}$$

$$= |I - iF'TF|^{-\frac{1}{2}} \exp \tfrac{1}{2}[\eta'\{iT - TF(I - iF'TF)^{-1}F'T\}\eta]$$

$$= |I - i\Sigma\, T|^{-\frac{1}{2}} \exp \tfrac{1}{2}\{\eta' T(I - i\Sigma\, T)^{-1}\eta\}, \tag{5.5}$$

where $T = \text{diag}\,(t_1, \ldots, t_k) \otimes I_n$.

Let $Z_{(j)} = \frac{1}{2}(SS_{(j)} - n - \eta_{(j)}' \eta_{(j)})/(n/2)^{\frac{1}{2}}$. Then

$$P(SS_{(1)}/2 \leq c^{-1} SS_{(j)}/2,\ j=2, \ldots, k) \tag{5.6}$$

$$= P\{Z_{(1)} \leq c^{-1} Z_{(j)} + (n/2)^{\frac{1}{2}}(c^{-1} - 1) + (\eta_{(j)}' \eta_{(j)}/c - \eta_{(1)}' \eta_{(1)})/(2n)^{\frac{1}{2}}, j=2, \ldots, k\}.$$

From the multivariate central limit theorem, it follows that for n large, the joint distribution of $Z_{(1)}, \ldots, Z_{(k)}$ does not depend upon $\eta_{(1)}, \ldots, \eta_{(k)}$. Moreover, since $\eta_{(1)}' \eta_{(1)} \leq \eta'_{(j)} \eta_{(j)}$, $j=2, \ldots, k$ by definition, the right hand side of (5.6) is greater than

$$P\{Z_{(1)} \le c^{-1}Z_{(j)} + (n/2)^{\frac{1}{2}}(c^{-1}-1) + \eta'_{(1)}\eta_{(1)}(c^{-1}-1)/(2n)^{\frac{1}{2}},\ j=2,\ldots,k\}$$

$$\ge P\{Z_{(1)} \le c^{-1}Z_{(j)} + (n/2)^{\frac{1}{2}}(c^{-1}-1)\}. \tag{5.7}$$

That is, the worst configuration (asymptotically) is when $\beta=0$. Bur now, the problem is the same as the one discussed in section 4, and thus the rule of (4.6) is appropriate for the present situation. Note that here $\Sigma = (\Sigma_{ij})$ where $\Sigma_{ij} = B_iB_j'$ is nxn as given in (5.4). Thus (4.2) becomes in this case

$$C_1=n(\Sigma_{j=1}^{k} t_j)/2,$$

$$C_2=n\{\Sigma_{j=1}^{k} t_j^2 + 2\Sigma_{i<j}t_it_j\mathrm{tr}(B_iB_j'B_jB_i')\}/2,$$

and

$$C_3=(2n/3)\{\Sigma_{j=1}^{k} t_j^3 + 3\Sigma_{i\neq j} t_i^2t_j\mathrm{tr}(B_iB_j'B_jB_i')$$

$$+6\Sigma_{h<i<j}t_ht_it_j\mathrm{tr}(B_hB_i'B_iB_jB_j'B_h)\}. \tag{5.8}$$

Expression (5.8) would determine an Edgeworth approximation of order $n^{-\frac{1}{2}}$. The remarks at the end of section 4 are again relevant. From a practical point of view, a user of the procedure outlined in (4.6) would want the selected subset to be small. Ideally, it would contain only the subset corresponding to the smallest sample residual sum of squares.

In McCabe and Arvesen [10], an algorithm for determining the parameter c for given P* and X using Monte Carlo methods is presented. An example with p=6 and t=3 is discussed in detail. A write-up for a FORTRAN program implementing this algorithm is available (McCabe, Arvesen and Pohl [11]). At the present time, an upper limit of 20 is set on the value of k. An increase in this limit would require additional storage and can be accomodated by changing certain dimension statements. For large k, the use of buffers may be necessary.

[1] Chambers, J. M., "On Methods of Asymptotic Approximation For Multivariate Distributions," Biometrika, 54 (1967), 367-83.

[2] Furnival, G. M., "All Possible Regressions With Less Computation," Technometrics, 13 (1971), 403-8.

[3] Gunst, R. F. and Webster, J. T., "Density Functions of the Bivariate Chi-square Distribution," Technical Report, Southern Methodist University, Dallas, Texas, (1972).

[4] Gupta, S. S. and Sobel, M., "On Selecting a Subset Containing the Population with the Smallest Variance," Biometrika, 49 (1962), 495-507.

[5] ——————, "On the Smallest of Several Correlated F Statistics," Biometrika, 49 (1962), 509-23.

[6] Hotelling, H., "The Selection of Variates for Use in Prediction with Some Comments on the Proglem of Nuisance Parameters," Annals of Mathematical Statistics, 11 (1940), 271-83.

[7] Jensen, D. R., "The Joint Distribution of Quadratic Forms and Related Distributions," Australian Journal of Statistics, 12 (1970), 13-22.

[8] Krishnamoorthy, A. S. and Parthasarathy, M. "A Multivariate Gamma-type Distribution," Annals of Mathematical Statistics 29 (1951), 719-36.

[9] La Motte, L. and Hocking, R. R., "Computational Efficiency in the Selection of Regression Variables," Technometrics, 12 (1970), 83-94.

[10] McCabe, G. P. and Arvesen, J. N., "A Subset Selection Procedure for Regression Variables," Journal of Statistical Computation and Simulation (to appear).

[11] McCabe, G. P., Arvesen, J. N., and Pohl, R. J., "A Computer Program for Subset Selection in Regression Analysis," Purdue University Department of Statistics Mimeo Series No. 317 (1973).

[12] Searle, S. R., Linear Models, New York: Wiley, 1972.

[13] Siotani, M., "The Extreme Value of the Generalized Distances of the Individual Points in the Multivariate Normal Sample," Annals of the Institute of Statistical Mathematics, 10 (1959), 183-206.

[14] Spjotvoll, E., "Multiple Comparison of Regression Functions," Annals of Mathematical Statistics, 43 (1972), 1076-88.