The Annals of Statistics
1983, Vol. 11, No. 1, 219–224

OPTIMAL PROPERTIES OF ONE-STEP VARIABLE SELECTION IN REGRESSION

BY RONALD W. BUTLER

The University of Texas

We consider the selection of the best subset of independent variables of a fixed size for possible inclusion in a regression model. The classical procedures (largest $\mathbb{R}^2$ to enter) are shown to be uniformly invariant Bayes in the sense of Paulson (1952) and Kudo (1956).

1. Introduction. This note is concerned with the problem of selecting the best subset of independent variables of a fixed size in a linear regression model. Under the usual assumptions of i.i.d. normal errors, Butler (1981) has shown that the natural selection procedure, which tests the significance of the subset with largest squared multiple correlation coefficient $\mathbb{R}^2$, is unique invariant Bayes and unique Bayes and hence admissible. In addition for the multivariate situation, he has also shown that the significance tests of the fixed size subsets which are most significant for either a Wilks' or Pillai's criterion have the same Bayes character. In Section 2 we strengthen these results in the case of the forward selection of a single independent variable by showing that the significance test to include the variable with largest partial correlation or, in the multivariate case, with largest partial first canonical correlation (largest Hotelling's T^2 for entry) uniformly minimizes the Bayes risk, given by the probability of incorrect selection, within a class of invariant selection rules. The method of proof follows the approach of Ferguson (1961) and Karlin and Truax (1960) who have considered the problem of single outlier detection. In Section 3 these results are extended to the selection of a single random effect variable in the multivariate case and a fixed number of random effect variables in the univariate case. Further examples of variable selection related to multiple outlier detection and discriminant analysis are given.

2. Selection of a single fixed effects variable. We denote

$$\underset{(n\times k)}{\tilde Y} = (\tilde y_1, \cdots, \tilde y_n)' = \underset{(n\times p)}{\tilde X} \cdot \underset{(p\times k)}{\tilde \Theta} + \underset{(n\times k)}{\tilde E} \tag{2.1}$$

as the null (H_0) model and use "~" to indicate a variable before transforming to canonical form. In (2.1) $\tilde X$ is fixed and $\tilde\Theta$ is fixed or random depending on the circumstances. Premultiplication of (2.1) by an $n \times n$ orthogonal matrix Ω having its first p rows as $(\tilde X' \tilde X)^{-1/2}\tilde X'$ transforms the model to canonical form which we indicate by removing the "~" in (2.1) so that $\Theta = (\tilde X'\tilde X)^{1/2}\,\tilde\Theta$ and $X' = (I, 0)$ where I is the $p \times p$ identity matrix. The rows of $\tilde E$ and hence E are taken as i.i.d. $MVN_k(0, \Sigma)$ with $\Sigma > 0$. Under this canonical form the least squares estimator (LSE) of Θ is $(y_1, \cdots, y_p)'$ and the matrix error sum of squares is

$$\hat S = (n-p)\hat\Sigma = \textstyle\sum_{i=p+1}^{n} y_i y_i' = \hat E'\hat E.$$

An idealized setting is assumed where $\tilde X$ is known to be relevant in the model and one other possibly relevant variable is included in an auxiliary list given by the columns of $n \times m$ matrix $\tilde Z = (\tilde z_1, \cdots, \tilde z_m)$. Let the $n \times m$ matrix $Z = \Omega(I - \tilde X(\tilde X'\tilde X)^{-1}\tilde X')\tilde Z$, so Z is the canonical form of $\tilde Z$ orthogonal to X (and the first p rows consist entirely of zeros). Also

Received May 1981; revised June 1982.
AMS 1970 *subject classifications.* Primary, 62C25, 62C10; secondary, 62J05, 62H30.
Key words and phrases. Variable selection, Bayesian decision rule, invariance, regression analysis.

assume that $z_i'z_i = 1\ \forall i$. Let H_i refer to the model

$$Y = X \cdot \Theta + z_i\tau_i' + E,$$

where τ_i is $k \times 1$. If $\tilde{X}$ contains a column of ones and $\{\tau_i\}$ are fixed effects, then $\Sigma^{-1/2}\tau_i$ is the vector of population (or long-run) partial correlations of the components of $\tilde{Y}$ with $\tilde{z}_i$ given $\tilde{X}$. Under H_i the LSE of Θ and τ_i are $(y_1, \cdots, y_p)'$ and $Y'z_i$ respectively.

The multidecision problem involved here compares H_0 and $\{H_i: i = 1, \cdots, m\}$ and a test for such a problem is an $(m + 1)$-vector $(\phi_0, \cdots, \phi_m)$ where ϕ_i is the probability of deciding H_i is the true model. If a zero-one loss function is assumed according to whether the decision is right-wrong, then the Bayes risk with respect to a prior weighting of the hypotheses is the probability of making an incorrect decision.

We follow the invariant Bayesian approaches of Ferguson (1961) and Karlin and Truax (1960), which represent refinements of the methods of Paulson (1952) and Kudo (1956), and restrict attention to the following class of tests:

(i) the test has level $0 < \alpha \leq 1$ so that

$$\mathscr{E}(\textstyle\sum_{i=1}^{m} \phi_i \mid H_0) = 1 - \mathscr{E}(\phi_0 \mid H_0) \leq \alpha.$$

(ii) the test is invariant to translation of $\tilde{Y}$ by a matrix of column vectors in the column space of $\tilde{X}$, thus allowing for the rescaling of $\tilde{X}$.

(iii) Invariance is also maintained under an arbitrary nonsingular $k \times k$ transformation of the rows of $\tilde{Y}$, thus allowing for the rescaling of $\tilde{Y}$.

(iv) Apriori, each hypothesis in $\{H_i: i = 1, \cdots, m\}$ is given the same weight $0 < \pi \leq m^{-1}$, where π is the value which assures the α level in (i).

We first consider the model I situation where $\{\tau_i\}$ are fixed effects and define the hypotheses for comparison as

$$H_0\colon \tau_i = 0\ \forall i, \quad H_i\colon \tau_i = \tau \neq 0 \quad \text{and} \quad \tau_j = 0\ \forall j \neq i \tag{2.2}$$

for $i = 1, \cdots, m$. This particular specification exhibits these characteristics.

(a) Hypothesis formulation (2.2) tests that the partial correlation vectors $\{\Sigma^{-1/2}\tau_i\}$ all vanish versus one vector has the value $\Sigma^{-1/2}\tau$. Because $Z'X = 0$, it is clear that the degree of orthogonality of each variable in $\tilde{Z}$ to design $\tilde{X}$ does not enter into the hypothesis test.

(b) Since

$$\mathscr{E}(\hat{S} \mid \Sigma, \tau_i, H_i) = (n - p)\Sigma + \tau_i\tau_i'$$

then apriori $\hat{S}$ has the same expected value under each H_i and becomes a useful benchmark with which to compare the effects of fitting the various $\{z_i\}$.

THEOREM 1. *Suppose $k < n - p$ and let $\hat{\rho}_i$ be the first* (*and only nonzero*) *partial canonical correlation of $\tilde{Y}$ and $\tilde{z}_i$ given $\tilde{X}$ so that*

$$\hat{\rho}_i^2 = z_i'\hat{E}\hat{S}^{-1}\hat{E}'z_i.$$

For hypothesis specification (2.2), *the test defined by*

$$\phi_{i^*} = 1 \quad \textit{if} \quad \hat{\rho}_{i^*}^2 = \max_i \hat{\rho}_i^2 > c_1$$

$$\phi_0 = 1 \quad \textit{otherwise}$$

minimizes the probability of an incorrect decision given by the Bayes risk with respect to (iv), *uniformly in $\tau \in \mathbb{R}^k$, among the class of tests specified in* (i)–(iv), *where $c_1 > 0$ is chosen so level α is maintained.*

PROOF. A test in the class above must be a function of the maximal invariant, so it suffices to consider its likelihood under each hypothesis. The residuals $\hat{E} = (0, \cdots, 0, y_{p+1}, \cdots, y_n)'$ are invariant to the transformations in (ii). Let $W = (y_{p+1}, \cdots, y_{p+k})$, which is $k \times k$ and nonsingular with probability one. To obtain invariance as in (iii) let

$$v_i = W^{-1}y_i, \quad i = 1, \cdots, n,$$

so that $(v_{p+1}, \cdots, v_{p+k}) = I_k$ and $v_1 = \cdots = v_p = 0$. By Lehmann (1959, Section 6.2, Theorem 2), $V = (v_1, \cdots, v_n) = W^{-1}\hat{E}'$ is the maximal invariant. Straightforward Jacobian transformations show that the likelihood of V is

$$L(V \mid H_i) \propto \| W \|^{n-p-k} \mid \Sigma \mid^{-(n-p)/2} \exp\{-\tfrac{1}{2}(V'W' - z_i\tau_i')\Sigma^{-1/2}(V'W' - z_i\tau_i')'\} \, dW.$$

Now let $S_0 = VV'$, $U = S_0^{1/2} W' \Sigma^{-1/2}$, and $N = S_0^{-1/2} V$ so that

$$\text{(2.3)} \qquad \begin{aligned} L(V \mid H_i) &\propto \mid S_0 \mid^{-(n-p)/2} \exp(-\tfrac{1}{2}\, \tau_i'\Sigma^{-1}\tau_i) \\ &\cdot \int \| U \|^{n-p-k} \cdot \operatorname{etr}\{-\tfrac{1}{2}(UU' - 2U'Nz_i\tau_i'\Sigma^{-1/2})\} \, dU. \end{aligned}$$

Now let $\Xi(\Gamma)$ be a $k \times k$ orthogonal matrix with first row given by $z_i'N'/\| z_i'N' \|$ $(\tau_i'\Sigma^{-1/2}/\| \tau_i'\Sigma^{-1/2} \|)$. Transform $\Psi = \Xi U \Gamma'$ and replace τ_i with τ so that

$$\text{(2.4)} \qquad \begin{aligned} L(V \mid H_i) &\propto \mid S_0 \mid^{-(n-p)/2} \exp(-\tfrac{1}{2}\tau'\Sigma^{-1}\tau) \\ &\cdot \int \| \Psi \|^{n-p-k} \exp\{-\tfrac{1}{2} \operatorname{tr} \Psi\Psi' + \psi_{11}(z_i'N'Nz_i)^{1/2}(\tau'\Sigma^{-1}\tau)^{1/2}\} \, d\Psi. \end{aligned}$$

As noted in Karlin and Truax (1960, page 320), expression (2.4) is an increasing function of

$$z_i'N'Nz_i = z_i'\hat{E}\hat{S}^{-1}\hat{E}'z_i = \hat{\rho}_i^2$$

for each value of $\tau'\Sigma^{-1}\tau$. The likelihood of H_0 is (2.4) with $\tau = 0$.

The Bayes risk with respect to the equally-weighted prior specification in (iv) is

$$\text{(2.5)} \qquad 1 - m\pi - \textstyle\sum_{i=1}^{m} \displaystyle\int \phi_i\{\pi L(V \mid \tau, H_i) - (1 - m\pi)L(V \mid H_0)\} \, dV$$

when we substitute $\phi_0 = 1 - \sum_{i=1}^{m} \phi_i$. The Bayes rule minimizing (2.5) accepts H_{i^*} if

$$L(V \mid \tau, H_{i^*}) = \max_i L(V \mid \tau, H_i) > \pi^{-1}(1 - m\tau)L(V \mid H_0)$$

and accepts H_0 otherwise. This is equivalent to the rule in Theorem 1 by the monotonicity of $L(V \mid \tau, H_i)$ in $\hat{\rho}_i^2$ for each value of τ. The minimization is also uniform in τ. □

EXAMPLE 1. Consider the classification of an unknown individual into one of two populations. Assume that $p = 1$ and $\tilde{X}$ is an $n \times 1$ column of ones. Let $(\tilde{y}_1, \cdots, \tilde{y}_\ell)$ and $(\tilde{y}_{\ell+1}, \cdots, \tilde{y}_{2\ell})$ represent observations from populations 1 and 2 respectively and suppose that observation $2\ell + 1 = n$ is to be classified. We let $m = 2$ and assume that τ_i is the differential effect of population 1 for $i = 1, 2$. If

$$\tilde{Z}' = \begin{pmatrix} \overbrace{1 \cdots 1}^{\ell} & \overbrace{-1 \cdots -1}^{\ell} & 1 \\ 1 \cdots 1 & -1 \cdots -1 & -1 \end{pmatrix},$$

then the selection of $\tilde{z}_i$ is equivalent to the classification of individual n into population i. We set $\alpha = 1$ $(\pi = \tfrac{1}{2})$ to eliminate the possible acceptance of H_0. From the variable updating formula

$$\hat{S}(X, z_i) = \hat{S}(X) - \hat{E}'z_iz_i'\hat{E},$$

where $\hat{S}(\cdot)$ is the matrix error sums of squares when fitting the enclosed variables, it follows that

$$\hat{\rho}_i^2 = 1 - \mid \hat{S}(X) \mid^{-1} \mid \hat{S}(X, z_i) \mid.$$

From this it is easy to show that

$$\hat{\rho}_i^2 = 1 - \frac{\mid \tilde{W} \mid \{1 + \ell(\ell + 1)^{-1}(\tilde{y}_n - \bar{y}_i)'\tilde{W}^{-1}(\tilde{y}_n - \bar{y}_i)\}}{\mid \tilde{T} \mid \{1 + (n - 1)n^{-1}(\tilde{y}_n - \bar{y})'\tilde{T}^{-1}(\tilde{y}_n - \bar{y})\}},$$

where $\bar{y}_i$ is the ith stratum mean without $\tilde{y}_n$, $\bar{y} = (\bar{y}_1 + \bar{y}_2)/2$, $\tilde{T}$ is the total sums of squares of the classified data $\tilde{y}_1, \cdots, \tilde{y}_{n-1}$ about its mean, and $\tilde{W}$ is the within strata sums of squares for the classified data. The sign of $\hat{\rho}_2^2 - \hat{\rho}_1^2$ is the same as the sign of $(\tilde{y}_n - \bar{y}_1)' \tilde{W}^{-1}(\tilde{y}_n - \bar{y}_1) - (\tilde{y}_n - \bar{y}_2)'\tilde{W}^{-1}(\tilde{y}_n - \bar{y}_2)$ which is the sign of $(\bar{y}_2 - \bar{y}_1)'\tilde{W}^{-1}(\tilde{y}_n - \bar{y})$, so Theorem 1 yields the plug-in version of Fisher's classification rule. The weaker admissible Bayes character of this test has been shown by Kiefer and Schwartz (1965) using a different approach.

3. Selection of random effect variables. We now assume $\{\tau_i\}$ represent random normal effects. In the multivariate situation we show the significance test on the single variable with largest partial first canonical correlation is again optimal. For the univariate case, the test of the variable subset of size $s \geq 1$ with largest partial multiple correlation (largest $\mathbb{R}^2$ to enter) among all subsets of size s is also optimal.

For brevity, we prove both of the above results within the notation of the multivariate model. The following notational extensions are necessary. Let $\mathscr{S}$ denote an arbitrary subset of $\{1, \cdots, m\}$ of size s, and, accordingly, suppose $\phi_{\mathscr{S}}$ is the probability of accepting $H_{\mathscr{S}}$, the hypothesis that the true model includes the additional variables in $\tilde{Z}_{\mathscr{S}} = (\tilde{z}_i: i \in \mathscr{S})$, which is $n \times s$. Let $\tau_{\mathscr{S}} = (\tau_i: i \in \mathscr{S})'$ be $s \times k$ and suppose in the univariate case $\hat{\rho}_{\mathscr{S}}$ is the partial multiple correlation of $\tilde{Y}$ and $\tilde{Z}_{\mathscr{S}}$ given $\tilde{X}$.

We again follow the invariant Bayesian approach of Section 2 and restrict attention to the tests of (i)–(iv) with the modification that equal prior weight is given to all $\binom{m}{s}$ possible subsets and $\mathscr{E}(\sum_{\mathscr{S}}\phi_{\mathscr{S}}| H_0) \leq \alpha$.

The hypotheses for comparison are

$$(3.1)\qquad \begin{aligned} &H_0: \tau_i = 0 \text{ w.p.1 } \forall i \\ &H_{\mathscr{S}}: (Z'_{\mathscr{S}}Z_{\mathscr{S}})^{1/2}\tau_{\mathscr{S}} \ \text{ has i.i.d. rows} \sim MVN_k(0, \lambda^2\Sigma) \quad \text{and} \quad \tau_i = 0 \text{ w.p.1 } \forall i \notin \mathscr{S}. \end{aligned}$$

Again $\mathscr{E}(\hat{S} \mid \Sigma, H_{\mathscr{S}}) = (n - p + s\lambda^2)\Sigma$ for all $\mathscr{S}$ so that $\hat{S}$ is a useful benchmark.

THEOREM 2. *For the univariate regression model* ($k = 1$) *consider the hypothesis specification in* (3.1) *with* $s \geq 1$, *and for the multivariate model assume* $s = 1$. *Then the rule*

$$(3.2)\qquad \begin{aligned} \phi_{\mathscr{S}^*} &= 1 \quad \textit{if} \quad \hat{\rho}^2_{\mathscr{S}^*} = \max_{\mathscr{S}}\hat{\rho}^2_{\mathscr{S}} > c_2 \\ \phi_0 &= 1 \quad \textit{otherwise} \end{aligned}$$

when $k = 1$ *and the rule of Theorem 1 when* $k > 1$ *minimize the probability of an incorrect decision given by the Bayes risk uniformly over* $\lambda > 0$ *among the class of tests in* (i)–(iv).

PROOF. The results here follow from the Bayes nature of the tests. Ferguson (1961) has shown that this result holds when looking for a single outlier in a multivariate location model.

We proceed as in Theorem 1 to (2.3) where it can be shown that the maximal invariant has the conditional distribution

$$(3.3)\qquad \begin{aligned} L(V \mid \tau_{\mathscr{S}}, H_{\mathscr{S}}) &\propto |S_0|^{-(n-p)/2}\exp(-\tfrac{1}{2} \operatorname{tr} \tau'_{\mathscr{S}}Z'_{\mathscr{S}}Z_{\mathscr{S}}\tau_{\mathscr{S}}\Sigma^{-1}) \\ &\int \| U\|^{n-p-k}\operatorname{etr}(-\tfrac{1}{2}\, UU' + U'NZ_{\mathscr{S}}\tau_{\mathscr{S}}\Sigma^{-1/2})\, dU. \end{aligned}$$

The marginal distribution of V under $H_{\mathscr{S}}$ follows by integrating (3.3) with respect to the random effect distribution of $\tau_{\mathscr{S}}$ in (3.1) so that

$$\begin{aligned} L(V \mid H_{\mathscr{S}}) \propto |S_0|^{-(n-p)/2}\lambda^{-sk} \int\int \| U\|^{n-p-k}\operatorname{etr}\{-\tfrac{1}{2}\, UU' - \tfrac{1}{2} f^{-2}\Lambda\Lambda' \\ + U'NZ_{\mathscr{S}}(Z'_{\mathscr{S}}Z_{\mathscr{S}})^{-1/2}\Lambda\}\, dU\, d\Lambda, \end{aligned}$$

where $\Lambda = (Z'_{\mathscr{S}}Z_{\mathscr{S}})^{1/2}\tau_{\mathscr{S}}\Sigma^{-1/2}$ and $f^2 = \lambda^2(1+\lambda^2)^{-1}$. Completing the square and integrating out Λ yields

$$(3.4)\qquad L(V\mid H_{\mathscr{S}}) \propto |S_0|^{-(n-p)/2}(1+\lambda^2)^{-ks/2}\int \|U\|^{n-p-k}\operatorname{etr}\{-\tfrac{1}{2}UU'(I-P)\}\,dU,$$

where $P = f^2NZ_{\mathscr{S}}(Z'_{\mathscr{S}}Z_{\mathscr{S}})^{-1}Z'_{\mathscr{S}}N'$. Make the substitution $A = UU'$ (which is not a square transformation but an integrated Jacobian of $|A|^{-1/2}$ is shown for it in Anderson, 1958, page 319) so that (3.4) takes the form of a Wishart integral and reduces to

$$(3.5)\qquad L(V\mid H_{\mathscr{S}}) \propto |S_0|^{-(n-p)/2}(1+\lambda^2)^{-ks/2}|I-P|^{-(n-p)/2}.$$

For both cases of Theorem 2, $|I-P| = 1 - f^2\hat{\rho}^2_{\mathscr{S}}$ since when $k = 1$, $P = f^2\hat{\rho}^2_{\mathscr{S}}$, and when $s = 1$,

$$Z'_{\mathscr{S}}N'\,NZ_{\mathscr{S}}(Z'_{\mathscr{S}}Z_{\mathscr{S}})^{-1} = \hat{\rho}^2_{\mathscr{S}}.$$

The Bayes risk with respect to our prior specification (iv) is

$$(3.6)\qquad \pi_0 - \textstyle\sum_{\mathscr{S}}\displaystyle\int \phi_{\mathscr{S}}[\pi L(V\mid\lambda, H_{\mathscr{S}}) - \pi_0 L(V\mid H_0)]\,dV$$

where $0 \le \pi_0 = 1 - \binom{m}{s}\pi < 1$ so that the rule

$$\phi_{\mathscr{S}^*} = 1 \quad \text{if} \quad L(V\mid\lambda, H_{\mathscr{S}^*}) = \max_{\mathscr{S}} L(V\mid\lambda, H_{\mathscr{S}}) > \pi^{-1}\pi_0 L(V\mid H_0)$$

$$\phi_0 = 1 \quad \text{otherwise}$$

is Bayes. Since (3.5) is a strictly increasing function of $\hat{\rho}^2_{\mathscr{S}}$ if either $k = 1$ or $s = 1$, then the tests of Theorem 2 are invariant Bayes for all $\lambda > 0$. □

EXAMPLE 2. Consider the detection of s outliers in a univariate location model having $\tilde{X}$ as an $n \times 1$ vector of ones. We suppose $m = n$, $\tilde{Z} = I_n$, and that the outliers result from independent normal random effects. Then $\mathscr{S} \subset \{1, \cdots, n\}$ indexes a possible subset of s outliers. Denote $\hat{S}(\tilde{X})$ as the sum of squares of the $\tilde{y}$-values about their mean, and $\hat{S}(\tilde{X}, \tilde{Z}_{\mathscr{S}})$ is the sum of squares of the $\tilde{y}$-values indexed by $\{1, \cdots, n\}\backslash\mathscr{S}$ about their mean. Then $\hat{\rho}^2_{\mathscr{S}} = 1 - \hat{S}(\tilde{X})^{-1}\hat{S}(\tilde{X}, \tilde{Z}_{\mathscr{S}})$ is the Grubb's (1950) outlier test statistic and its uniformly invariant Bayes character is given in Theorem 2.

EXAMPLE 3. Consider the classification of an individual into one of $m + 1$ populations based on a single variable ($k = 1$). Suppose that $\tilde{y}_{(i-1)\ell+1}, \cdots, \tilde{y}_{i\ell}$ are derived from population i for $i = 1, \cdots m + 1$, and observation $(m+1)\ell + 1 = n$ is from the unclassified individual. Random effect interpretations are given to the strata effects here which must sum to zero w.p.1 if $\tilde{X}$ is taken as a column of ones. Let $\tilde{z}_i$, $\tilde{z}_i^*$ and $\tilde{z}_i^{**}$ respectively be the usual $n \times 1$ design variables indicating stratum i effects with a 0, 1 and -1 in the nth row for $i = 1, \cdots, m$. Then $\tilde{Z}$ consists of these $3m$ variables and we must select from

$$\mathscr{S}_1 = \{\tilde{z}_1^*, \tilde{z}_2, \cdots, \tilde{z}_m\}, \cdots, \mathscr{S}_m = \{\tilde{z}_1, \cdots, \tilde{z}_{m-1}, \tilde{z}_m^*\}, \quad \text{and} \quad \mathscr{S}_{m+1} = \{\tilde{z}_1^{**}, \cdots, \tilde{z}_m^{**}\},$$

where the selection of $\mathscr{S}_i$ corresponds to classification into population i. Then, in the notation of Example 1 extended to $m + 1$ strata,

$$\hat{\rho}^2_{\mathscr{S}_i} = 1 - \frac{\tilde{W} + \ell(\ell+1)^{-1}(\tilde{y}_n - \bar{y}_i)^2}{\tilde{T} + (n-1)n^{-1}(\tilde{y}_n - \bar{y})^2}$$

and $\hat{\rho}^2_{\mathscr{S}_i}$ is largest at the i for which $(\tilde{y}_n - \bar{y}_i)^2$ is smallest.

Acknowledgment. The careful reading of a referee and his detailed comments have greatly improved the paper and the author is very grateful.

REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

BUTLER, R. W. (1981). The admissible Bayes character of subset selection techniques involved in variable selection, outlier detection and slippage problems. *Ann. Statist.* **9** 960–973.

FERGUSON, T. S. (1961). On the rejection of outliers. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 253–287.

GRUBBS, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21** 27–58.

KARLIN, S. and TRUAX, D. R. (1960). Slippage problems. *Ann. Math. Statist.* **31** 296–324.

KIEFER, J. and SCHWARTZ, R. (1965). Admissible Bayes character of T^2-, R^2-, and other fully invariant tests for classical multivariate normal problems. *Ann. Math. Statist.* **36** 747–770.

KUDO, A. (1956). On the testing of outlying observations. *Sankhyā Ser. A* **17** 67–76.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

PAULSON, E. (1952). An optimum solution to the K-sample slippage problem for the normal distribution. *Ann. Math. Statist.* **23** 610–616.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712