NON-OPTIMALITY OF PRELIMINARY-TEST ESTIMATORS FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION[1]

BY STANLEY L. SCLOVE,[2] CARL MORRIS

AND R. RADHAKRISHNAN[3]

Carnegie-Mellon University and RAND

Estimation-preceded-by-testing is studied in the context of estimating the mean vector of a multivariate normal distribution with quadratic loss. It is shown that although there are parameter values for which the risk of a preliminary-test estimator is less than that of the usual estimator, there are also values for which its risk exceeds that of the usual estimator, and that it is dominated by the positive-part version of the Stein-James estimator. The results apply to preliminary-test estimators corresponding to any linear hypothesis concerning the mean vector, e.g., an hypothesis in a regression model. The case in which the covariance matrix of the multinormal distribution is known up to a multiplicative constant and the case in which it is completely unknown are treated.

1. Introduction and summary. In making statistical inferences relative to incompletely specified models, we sometimes test an hypothesis concerning some or all of the parameters before performing an estimation, making the estimation procedure dependent upon the outcome of the test of hypothesis. For example, in multiple regression we may test the significance of the regression upon a subset of variables and include or exclude these variables from the model, depending upon the outcome of the test of significance. The estimate of the coefficients of the remaining variables in general depends, then, upon the outcome of the test of significance. We shall consider such procedures within the following context. Let X be a p-variate ($p \geq 3$) normal random vector with unknown mean vector θ and convariance matrix $\sigma^2 I_p$, σ^2 being unknown. (Here I_p denotes the identity matrix of order p.) Let S be distributed independently of X as $\sigma^2 \chi_n{}^2$. ($\chi_n{}^2$ denotes a chi-square random variable with n degrees of freedom.) We shall consider the problem of estimating θ when the loss function is

$$(1.1) \qquad L(\hat{\theta}; \theta, \sigma^2) = ||\hat{\theta} - \theta||^2/\sigma^2 ,$$

where, for a vector v, $||v||^2 \equiv v'v$. The risk of an estimator $\varphi(X, S)$ is thus $E[||\varphi(X, S) - \theta||^2]/\sigma^2 = R(\varphi; \theta, \sigma^2)$, say. The usual estimator, $\varphi(X, S) = X$, has constant risk, equal to p for all θ, σ^2; it is well known that X is unbiased, is the maximum likelihood estimator, and is minimax. Stein and James [2] showed

---

that X is dominated by the estimator

(1.2) $\varphi(X, S) = (1 - cS/X'X)X$,   where $0 < c < 2(p - 2)/(n + 2)$.

Consider the following procedure of testing-followed-by-estimation. We test the hypothesis H: $\theta = 0_p$ against not H (0_p denotes the vector of p zeros), using the statistic $F = X'X/S$ and rejecting H if $F > c > 0$; note that under H the statistic nF/p is distributed according to Snedecor's F-distribution with p and n degrees of freedom. If we accept H we take 0_p as our estimate of θ; otherwise, we use the usual estimate, X. The resulting "preliminary-test estimator" can be written as

(1.3) $\varphi(X, S) = I_{(c, \infty)}(F)X$,

where $I_B(\cdot)$ denotes the indicator function of the set B. In Section 2 expressions for the risk of invariant estimators of θ are given. In Section 3 these formulas are used to show that there are parameter values for which the risk of (1.3) exceeds p, that there are values for which its risk is less than p, and that (1.3) is dominated by a modification of (1.2) which is minimax if the size of the test $\theta = 0_p$ is not too small. It is shown in Section 4 that the results extend immediately to preliminary-test estimators corresponding to other hypotheses concerning the mean vector, e.g., the hypothesis that it lies in a specified linear subspace. In Section 5 covariance structures other than $\sigma^2 I_p$ are considered; in particular it is shown that results for the case $\sigma^2 I_p$ extend at once to the case in which the covariance matrix is completely unknown.

2. Expressions for the risk of invariant estimators. We shall be concerned with estimators

$$\varphi_h(X, S) = h(F)X ,$$

where $h(\cdot)$ is any Borel-measurable function on $[0, \infty)$. Such estimators are invariant estimators under the transformation $X \to cPX$, $S \to c^2S$, where c is any scalar and P is any $p \times p$ orthogonal matrix ($P'P = I$).

(An example of the model under consideration arises as follows. Let X_1, X_2, $\cdots$, X_n be independent and identically distributed according to a p-variate normal distribution with mean vector θ and covariance matrix $\tau^2 I$. Then $X = \bar{X} = \sum_{i=1}^{n} X_i/N$ is p-variate normal with mean vector θ and covariance matrix $\sigma^2 I$, where $\sigma^2 = \tau^2/N$, and $S = \sum_{j=1}^{p} \sum_{i=1}^{N} (X_{ij} - \bar{X}_j)^2$, where $\bar{X}_j$ and X_{ij} are the jth components of $\bar{X}$ and X_i, respectively, is distributed independently of X as $\sigma^2 \chi_n^2$, where $n = p(N - 1)$. If we make the transformation $X_i \to cPX_i$, then the induced transformation of X and S is $X \to cPX$, $S \to c^2S$.)

LEMMA 1.

(2.1) $R(\varphi_h; \theta, \sigma^2) = E[h^2(\chi_{p+2K}^2/\chi_n^2)\chi_{p+2K}^2] - 4E[h(\chi_{p+2K}^2/\chi_n^2)K] + \theta'\theta/\sigma^2$,

where K is distributed according to the Poisson distribution with parameter $\theta'\theta/2\sigma^2$ and χ_{p+2K}^2 and χ_n^2 are independent.

PROOF. (cf. formula (2.11) of [3].)

LEMMA 2. *If $h(\cdot)$ is any Borel-measurable function on $[0, \infty)$, then*

$$(2.2) \qquad E[h(\chi^2_{p+2K}/\chi_n^2)K] = (\theta'\theta/2\sigma^2)E[\chi^2_{p+2+2K}/\chi_n^2)]$$

and

$$(2.3) \qquad E[h(\chi^2_{p+2K}/\chi_n^2)\chi^2_{p+2K}] = pE[h(\chi^2_{p+2+2K}/\chi_n^2)] + (\theta'\theta/\sigma^2)E(\chi^2_{p+4+2K}/\chi_n^2)],$$

where χ^2_{p+2K} and χ_n^2 are independent, χ^2_{p+2+2K} and χ_n^2 are independent, and χ^2_{p+4+2K} and χ_n^2 are independent.

PROOF.

$$E[h(\chi^2_{p+2K}/\chi_n^2)K] = \sum_{k=0}^{\infty} k \frac{(\theta'\theta/2\sigma^2)^k}{k!} \exp[-\theta'\theta/2\sigma^2]E[h(\chi^2_{p+2K}/\chi_n^2) \mid K = k]$$

$$(2.4) \qquad = \frac{\theta'\theta}{2\sigma^2} \sum_{k=1}^{\infty} \frac{(\theta'\theta/2\sigma^2)^{k-1}}{(k-1)!} \exp[-\theta'\theta/2\sigma^2]E[h(\chi^2_{p+2k}/\chi_n^2)]$$

$$= \frac{\theta'\theta}{2\sigma^2} \sum_{l=0}^{\infty} \frac{(\theta'\theta/2\sigma^2)^l}{l!} \exp[-\theta'\theta/2\sigma^2]E[h(\chi^2_{p+2+2l}/\chi_n^2)]$$

$$= (\theta'\theta/2\sigma^2)E[h(\chi^2_{p+2+2K}/\chi_n^2)],$$

establishing (2.2). The density of χ_p^2 is $f_p(s) = s^{p/2-1}e^{-s/2}/\Gamma(p/2)2^{p/2}$. Note that $sf_p(s) = pf_{p+2}(s)$. This gives $E[h(\chi^2_{p+2k}/\chi_n^2)\chi^2_{p+2k}] = (p + 2k)E[h(\chi^2_{p+2k+2}/\chi_n^2)]$, whereupon $E[h(\chi^2_{p+2K}/\chi_n^2)\chi^2_{p+2K}] = E[(p + 2K)h(\chi^2_{p+2K+2}/\chi_n^2)]$. (2.3) follows from this relation and (2.2).

LEMMA 3.

$$(2.5) \qquad R(\varphi_h; \theta, \sigma^2) = E[h^2(\chi^2_{p+2K}/\chi_n^2)\chi^2_{p+2K}] - 2(\theta'\theta/\sigma^2)E[h(\chi^2_{p+2+2K}/\chi_n^2)]$$
$$+ \theta'\theta/\sigma^2$$

$$(2.6) \qquad = pE[h^2(\chi^2_{p+2+2K}/\chi_n^2)] - 2(\theta'\theta/\sigma^2)E[h(\chi^2_{p+2+2K}/\chi_n^2)]$$
$$+ (\theta'\theta/\sigma^2)E[h^2(\chi^2_{p+4+2K}/\chi_n^2)] + \theta'\theta/\sigma^2.$$

PROOF. Use of (2.2) in (2.1) produces (2.5), and use of (2.3) in (2.5) gives (2.6).

3. Properties of preliminary-test estimators.

THEOREM 1. *Let $\varphi(X, S)$ be of the form (1.3), where c is any positive constant. Then $R(\varphi; \theta, \sigma^2) > p$ if $\|\theta\|^2/\sigma^2 > p$.*

Since the risk of X is p, and X is minimax, Theorem 1 states that no such preliminary-test estimator can be minimax.

PROOF. Here $h(F) = I_{(c,\infty)}(F)$, so that $h^2(F) = h(F)$, and the risk is thus

$$(3.1) \qquad R(\varphi; \theta, \sigma^2) = (p - 2\theta'\theta/\sigma^2) \Pr\{\chi^2_{p+2+2K} > c\chi_n^2\}$$
$$+ (\theta'\theta/\sigma^2) \Pr\{\chi^2_{p+4+2K} > c\chi_n^2\} + \theta'\theta/\sigma^2$$
$$\geqq (p - 2\theta'\theta/\sigma^2) \Pr\{\chi^2_{p+2+2K} > c\chi_n^2\}$$

$$(3.2) \qquad + (\theta'\theta/\sigma^2) \Pr\{\chi^2_{p+2+2K} > c\chi_n^2\} + \theta'\theta/\sigma^2$$
$$= (p - \theta'\theta/\sigma^2) \Pr\{\chi^2_{p+2+2K} > c\chi_n^2\} + \theta'\theta/\sigma^2$$
$$= p + (\theta'\theta/\sigma^2 - p) \Pr\{\chi^2_{p+2+2K} \leqq c\chi_n^2\},$$

where (3.2) follows because χ^2_{p+4+2K} is stochastically larger than χ^2_{p+2+2K}. The result now follows, because the final expression is greater than p if $\theta'\theta/\sigma^2 > p$.

Theorem 1 states that there are points in the parameter space where the preliminary-test estimator has larger risk than does the usual estimator, X; we would expect preliminary-test estimators to perform better than X when the length of θ is small; we see in the next theorem that this is true.

THEOREM 2.   Let $\varphi(X, S)$ be of the form (1.3).   Then $R(\varphi; \theta, \sigma^2) < p$ if $\theta'\theta/\sigma^2 < p/2$.

PROOF.   Choose θ, σ^2 such that $p - 2\theta'\theta/\sigma^2 > 0$. Replacing $\Pr\{\chi^2_{p+2+2K} > c\chi_n{}^2\}$ in (3.1) by $\Pr\{\chi^2_{p+4+2K} > c\chi_n{}^2\}$, which is a larger probability, we see that

$$
\begin{aligned}
R(\varphi; \theta, \sigma^2) &\leqq (p - 2\theta'\theta/\sigma^2)\Pr\{\chi^2_{p+4+2K} > c\chi_n{}^2\} \\
&\quad + (\theta'\theta/\sigma^2)\Pr\{\chi^2_{p+4+2K} > c\chi_n{}^2\} + \theta'\theta/\sigma^2 \\
&= (p - \theta'\theta/\sigma^2)\Pr\{\chi^2_{p+4+2K} > c\chi_n{}^2\} + \theta'\theta/\sigma^2 \\
&= p - (p - \theta'\theta/\sigma^2)\Pr\{\chi^2_{p+4+2K} \leqq c\chi_n{}^2\} \\
&< p,
\end{aligned}
$$

since $\theta'\theta/\sigma^2 < 2\theta'\theta/\sigma^2 < p$.

The risk of preliminary-test estimators at the origin can readily be computed from Lemma 1. For when $\theta = 0_p$, K is degenerate at 0 and $F = \chi_p{}^2/\chi_n{}^2$, and so $R(\varphi; 0_p, \sigma^2) = E[I_{(c,\infty)}(\chi_p{}^2/\chi_n{}^2)\chi_p{}^2]$. As is well known, the statistic $G = \chi_p{}^2 + \chi_n{}^2$ is independent of F, and we have $\chi_p{}^2 = GF/(1 + F)$, so

$$
\begin{aligned}
R(\varphi; 0_p, \sigma^2) &= E[I_{(c,\infty)}(F)GF/(1 + F)] \\
&= (n + p)E[I_{(c,\infty)}(F)F/(1 + F)] \\
&= (n + p)E[I_{(c/(1+c),\infty)}(B)B] \\
&= [(n + p)/\beta(p/2, n/2)]\int_{c/(1+c)}^{\infty}b^{p/2}(1 - b)^{n/2-1}db \\
&= (n + p)\beta(p/2 + 1, n/2)\Pr\{B^* \geqq c/(1 + c)\}/\beta(p/2, n/2) \\
&= p\Pr\{B^* \geqq c/(1 + c)\} \\
&= p\Pr\{\chi^2_{p+2}/\chi_n{}^2 \geqq c\},
\end{aligned}
$$

where $B \equiv F/(1 + F)$ has the beta distribution with parameters $p/2$ and $n/2$ and B^* has the beta distribution with parameters $(p + 2)/2$ and $n/2$. Thus, as $c \to \infty$, the risk at the origin of the corresponding preliminary-test estimators tends to zero, that is, we can find a preliminary-test estimator which does arbitrarily well at the origin (at the expense of being poor elsewhere).

We turn now to the main result of this paper, that any preliminary test estimator is dominated by a modification of Stein's rule (1.2) if $p \geqq 3$. The preliminary test rule φ_0 estimates θ as 0 if $F \leqq c$ and as X otherwise. Since Stein's rules dominate X, we expect that the rule which estimates θ as 0_p if $F \leqq c$ and otherwise estimates θ by a Stein rule would be superior. That this is true is the subject of Theorem 3. Define $c_0 \equiv (p - 2)/(n + 2)$ and let a satisfy $0 < a \leqq 2$.

Define $g_0(F) = I_{(c,\infty)}(F)$, $g_1(F) = g_0(F)[1 - ac_0/F]$ and $g_2(F) = I_{(ac_0,\infty)}(F)[1 - ac_0/F]$. Define $\varphi_i(X, S) = g_i(F)X$, $i \leq 0, 1, 2$. Theorem 3 shows φ_1 dominates φ_0 for all c, a. If $c \leq ac_0$ then φ_2 is uniformly better than φ_1 since $g_2(F) = \max(0, g_1(F))$ is a uniformly better coefficient of X than $g_1(F)$ in that g_2 does not change the signs of the components of X. (This is pointed out in [3].) The estimators φ_2 are always minimax and therefore the rule φ_2 is a minimax substitute for the preliminary test estimator whenever $c \leq 2c_0$ (by choosing a to satisfy $c/c_0 \leq a \leq 2$ in g_2). If $c > 2c_0$ then φ_1 dominates φ_0, but φ_2 may not dominate φ_0. In this case, φ_1 may not be minimax. All of these assertions are simple consequences of the following

THEOREM 3. *Let $c_0 = (p - 2)/(n + 2)$, let a satisfy $0 < a \leq 2$, and define $g_0(F) \equiv I_{(c,\infty)}(F)$, $g_1(F) \equiv g_0(F)[1 - ac_0/F]$, $\varphi_i(X, S) \equiv g_i(F)X$, $i = 0, 1$. Then $R(\varphi_0; \theta, \sigma^2) > R(\varphi_1; \theta, \sigma^2)$ for all θ, σ^2.*

PROOF. In formula (2.1), we calculate risks conditioned on the value $K = k$ and set $F = \chi^2_{p+2K}/\chi^2_n$ which is independent of $T = \chi^2_{p+2K} + \chi^2_n$. Thus, $Eh^2(\chi^2_{p+2K}/\chi^2_n)\chi^2_{p+2K} = Eh^2(F)TF/(1 + F) = (n + p + 2k)Eh^2(F)F/(1 + F)$. Recall that the density of the ratio χ^2_u/χ^2_v at f is $f^{u/2-1}/[(1 - f)^{(u+v)/2}\beta(u/2, v/2)]$. Let F_1 have the distribution of $\chi^2_{p+2K}/\chi^2_{n+2}$. Then for any function h, $Eh(F)/(1 + F) = n(p + 2k + n)^{-1}Eh(F_1)$. Formula (2.1) therefore becomes

$$(3.3) \qquad R(\varphi_h; \theta, \sigma^2) = \sum_{k=0}^{\infty} \frac{\lambda^k \exp(-\lambda)}{k!} R_k(h)$$

with $\lambda \equiv \theta'\theta/2\sigma^2$ and

$$(3.4) \qquad R_k(h) \equiv nEh^2(F_1)F_1 - \frac{4nk}{p + 2k + n} Eh(F_1)(1 + F_1) + 2\lambda .$$

Since $g_0^2(F) = g_0(F)$ and $g_1^2(F) = g_0(F)[1 - 2ac_0/F + a^2c_0^2/F^2]$, from (3.4) we easily derive

$$(3.5) \qquad \Delta R_k \equiv R_k(g_0) - R_k(g_1)$$
$$= nac_0 Eg_0(F_1)[2 - ac_0/F_1] - \frac{4nkac_0}{n + p + 2k} Eg_0(F_1)\frac{1 + F_1}{F_1} .$$

Let F_2 and F_3 have the distributions of $\chi^2_{p-2+2K}/\chi^2_{n+2}$ and $\chi^2_{p-2+2K}/\chi^2_{n+4}$ respectively. Then it is easily checked by writing out the densities of F_1, F_2, F_3 that for any function h,

$$Eh(F_1)(1 + F_1)/F_1 = \frac{n + p + 2k}{p - 2 + 2k} Eh(F_2)$$

$$Eh(F_1)/F_1 = \frac{n + 2}{p - 2 + 2k} Eh(F_3) .$$

Consequently,

$$\Delta R_k/nac_0 = 2Eg_0(F_1) - \frac{ac_0(n + 2)}{p - 2 + 2k} Eg_0(F_3)$$

$$- 2\frac{2k}{p - 2 + 2k} Eg_0(F_2) ,$$

or

$$(3.6) \qquad \Delta R_k / nac_0 = 2E[g_0(F_1) - g_0(F_2)] + \frac{p - 2}{p - 2 + 2k} [2Eg_0(F_2) - aEg_0(F_3)] .$$

From their definitions, F_1, F_2, F_3 are stochastically ordered, satisfying $F_1 > F_2 > F_3$. Since g_0 is monotone increasing and $2 \geqq a$, we have the bound

$$(3.7) \qquad \Delta R_k > \frac{nac_0(p - 2)(2 - a)}{p - 2 + 2k} \Pr\{F_2 \geqq c\} \geqq 0 .$$

We therefore have

$$(3.8) \qquad R(\varphi_0; \varphi, \sigma^2) - R(\varphi_1; \theta, \sigma^2) = \sum_{k=0}^{\infty} \frac{\lambda^k \exp(-\lambda)}{k!} \Delta R_k > 0 ,$$

which proves the theorem.

Averaging ΔR_k of (3.6) according to the formula (3.3) provides an exact expression for $R(\varphi_0; \theta, \sigma^2) - R(\varphi_1; \theta, \sigma^2)$. The risk of $\varphi_0(X, S)$ may be calculated from (3.1). The difference of these two results provides a formula for $R(\varphi_1; \theta, \sigma^2)$ in terms of the distribution functions of noncentral F distributions.

A limiting case (corresponding to $n = \infty$) is essentially the special case of σ^2 known. Here also a result similar to that of Theorem 3 holds. From calculations similar to, but easier than those of Theorem 3, using the risk formula (2.10) of [3] page 353, it is easily shown that the estimator $I_{(c,\infty)}(V)X$, $V \equiv ||X||^2$, is dominated by Stein's modified estimator $I_{(c,\infty)}(V)[1 - a(p - 2)\sigma^2/V]X$ for any $0 < a \leqq 2$, $c > 0$. Again, Stein's modified estimator above can be uniformly improved upon by its positive part if $c < a(p - 2)$, and will not be minimax for large c.

We have shown that minimax Stein estimators of the form $\varphi_2(X, S)$ are uniformly better than preliminary-test estimators provided $c \leqq 2c_0$, or equivalently, provided the size α of the test that $\theta = 0$ is not too small. Given in Table I for several values of n, p are the probabilities $\alpha = \Pr\{\varphi_0(X, S) \neq 0\}$ assuming $\theta = 0$ for $c = c_0$ and $c = 2c_0$. Letting $F_{p,n}$ be Snedecor's F variable, α is calculated as

$$\alpha = \Pr\{F_{p,n} > nac_0/p\}$$

for $a = 1, 2$ in Table I. As $n \to \infty$, the case of σ known, the value calculated is

TABLE I

Levels of significance of tests corresponding to use of the estimator
$I_{(c,\infty)}(F)X$ *when* $c = c_0$ *and* $c = 2c_0$, *where* $c_0 = (p - 2)/(n + 2)$

		p					
		3	4	5	10	20	30	∞
$n = p$	c_0	.89	.84	.81	.73	.67	.64	.50
$n = p$	$2c_0$	.76	.65	.57	.33	.14	.07	0
$n = \infty$	c_0	.80	.74	.70	.63	.59	.57	.50
$n = \infty$	$2c_0$	.57	.41	.31	.10	.02	.003	0

$\alpha = \Pr\{\chi_p{}^2 > a(p - 2)\}$.   From the table, we see that for small values of p, minimax Stein estimators only dominate those preliminary-test estimators with larger values of α than are ordinarily used. The values in the table are not the minimum possible values for which a minimax substitute exists dominating the preliminary test estimator: the Stein rules are inadmissible; and rules of the form $\varphi_1(X, S)$ may be minimax for some values of $c > 2c_0$.

4. Other hypotheses. The results extend immediately to preliminary-test and positive-part estimators corresponding to hypotheses more complicated than the hypothesis that the mean vector equals the zero vector, for this hypothesis is essentially the canonical form of more complicated hypotheses.

The hypothesis that part of the mean vector is equal to a given vector. Suppose that the random vector Z is distributed according to a q-variate normal distribution with mean vector ζ and covariance matrix $\sigma^2 I_q$, where Z and ζ are partitioned as $Z' = (Z_1', Z_2')$, $\zeta' = (\zeta_1', \zeta_2')$, Z_1 and ζ_1 being $p \times 1$ and Z_2 and ζ_2 being $q - p \times 1$. S is distributed independently of Z as $\sigma^2 \chi_n{}^2$. Suppose we are interested in the hypothesis $\zeta_1 = \zeta_{10}$. By taking $\theta = \zeta_1 - \zeta_{10}$, $X = Z_1 - \zeta_{10}$, and $F = ||Z_1 - \zeta_{10}||^2/S$ we can see that for the problem of estimating ζ subject to . loss $||\hat{\zeta} - \zeta||^2/\sigma^2 (= ||\hat{\zeta}_1 - \zeta_1||^2/\sigma^2 + ||\hat{\zeta}_2 - \zeta_2||^2/\sigma^2)$ the estimator

$$(4.1) \qquad \begin{pmatrix} \left(1 - cS/||Z_1 - \zeta_{10}||^2\right)^+ (Z_1 - \zeta_{10}) + \zeta_{10} \\ Z_2 \end{pmatrix}$$

dominates the estimator

$$\begin{pmatrix} I_{(c,\infty)}(F)(Z_1 - \zeta_{10}) + \zeta_{10} \\ Z_2 \end{pmatrix}.$$

A special case is the case $p = q$, when Z_1 is the whole vector Z.

The hypothesis that the mean vector lies in a specified linear subspace. Let T be a t-dimensional normal random vector with mean vector τ and covariance matrix $\sigma^2 I_t$. Let S be distributed independently of T as $\sigma^2 \chi_n{}^2$. We are interested in the hypothesis that τ lies in ω, a specified linear subspace of dimension $t - p$. (Thus p is the number of degrees of freedom associated with the hypothesis.) Let $\hat{T}$ be the projection of T onto ω, i.e., $\hat{T} = PT$, where P is the appropriate projection matrix. A preliminary test estimator corresponding to the hypothesis takes the value $\hat{T}$ or T, according as we accept or reject the hypothesis. Thus a preliminary-test estimator for this problem takes the form $I_{(c,\infty)}(F)(T - \hat{T}) + \hat{T}$, i.e.,

$$(4.2) \qquad I_{(c,\infty)}(F)(I - P)T + PT,$$

where here the test statistic $F = ||(I - P)T||^2/S$. In the canonical reduction corresponding to the hypothesis $\tau \in \omega$ we make an orthogonal transformation of T into the vector Z, the transformation being chosen so that the hypothesis $\tau \in \omega$ is transformed into the hypothesis $\zeta_1 = 0_p$. It can be shown that in terms of the vector T the estimator (4.1) takes the form

$$(4.3) \qquad \left(1 - \frac{cS}{||(I - P)T||^2}\right)^+ (I - P)T + PT.$$

It follows that for the problem of estimating τ, subject to loss $||\hat{\tau} - \tau||^2/\sigma^2$, the positive-part estimator (4.3) dominates the preliminary-test estimator (4.2).

A particularly interesting special case is the hypothesis that all the components of τ are equal. The linear subspace ω is then the equi-angular line in t-space, and its dimensionality, $t - p$, is equal to one, i.e., $p = t - 1$, so that the results require $t \geqq 4$. The projection $\hat{T}$ is in this case equal to $\bar{T}e_t$, where e_t is the vector of t ones and $\bar{T} = \sum_{i=1}^{t} T_i/t$, T_i being the ith component of T. It follows that (4.3) is

$$\left(1 - \frac{cS}{\sum_{i=1}^{t} (T_i - \bar{T})^2}\right)^+ (T - \bar{T}e_t) + \bar{T}e_t \,.$$

Regression models. Let V be distributed according to an N-variate normal distribution with covariance matrix $\sigma^2 I_N$ and mean vector ν lying in Ω, an r-dimensional linear subspace of N-space $(r \leqq N)$. (Thus ν is of the form $\nu = A\beta$, where the parameter vector β ranges over k-space and A is an $N \times k$ matrix of constants and has rank $r \leqq k$.) A linear hypothesis concerning ν states that ν lies in ω, an $(r - p)$-dimensional subspace of $\Omega(p < r)$. Let P_Ω be the matrix giving the projection from N-space to Ω; $\hat{V}_\Omega = P_\Omega V$ is the matrix projection of V onto Ω. The statistic $S = ||(I - P_\Omega)V||^2$ is distributed independently of $\hat{V}$ as $\sigma^2 \chi^2_{N-r}$. Let P_ω be the matrix giving the projection onto ω; $\hat{V}_\omega = P_\omega V$ is the projection of V onto ω. The appropriate preliminary-test estimator estimates ν as $\hat{V}_\omega$ if we accept the hypothesis that $\nu \in \omega$ and as $\hat{V}_\Omega$ if we reject this hypothesis and thus takes the form $I_{(c,\infty)}(F)(\hat{V}_\Omega - \hat{V}_\omega) + \hat{V}_\omega$, i. e.,

(4.4) $I_{(c,\infty)}(F)(P_\Omega - P_\omega)V + P_\omega V \,,$

where here the test statistic is $F = ||(P_\Omega - P_\omega)V||^2/S$. In the canonical reduction of this problem we transform to a vector U with mean vector μ. These vectors are partitioned as $U' = (U_1', U_2', U_3')$ and $\mu' = (\mu_1', \mu_2', \mu_3')$, where U_1 and μ_1 are $p \times 1$, U_2 and μ_2 are $(N - p + r) \times 1$, and U_3 and μ_3 are $(N - r) \times 1$. The problem of testing the hypothesis that $\nu \in \omega$ is transformed into testing $\mu_1 = 0_p$ against $\mu_1 \neq 0_p(\mu_3 = 0_{N-r}$ under both the null and alternative hypotheses and μ_2 is unspecified). Our results are applied to this model by taking $\zeta_1 = \mu_1$, $\zeta_2 = \mu_2$, $Z_1 = U_1$, $Z_2 = U_2$. In terms of V and $S = ||(I - P_\Omega)V||^2$ the resulting positive-part estimator takes the form

(4.5) $\left(1 - \frac{cS}{||(P_\Omega - P_\omega)V||^2}\right)^+ (P_\Omega - P_\omega)V + P_\omega V \,.$

The result is that for the problem of estimating ν subject to loss $||\hat{\nu} - \nu||^2/\sigma^2$ the positive-part estimator (4.5) dominates the preliminary-test estimator (4.4).

By referring to [3], where it is shown how to apply the results of [2], to analysis-of-variance problems (by projection onto appropriate linear subspaces, as above), one can generate positive-part estimators which dominate the corresponding preliminary-test estimators for these problems, too.

5. Other covariance structures. Let Σ be the covariance matrix of the multinormal random vector X. We assumed that $\Sigma = \sigma^2 I$, where σ^2 is unknown but S, independent of X and distributed as $\sigma^2 \chi_n^2$, is available for estimating σ^2. The results obtained extend at once to other covariance structures.

$\Sigma = \sigma^2 B$. Here B is a known, symmetric, positive-definite matrix. This case is treated by making a transformation which sends B into I.

$\Sigma = \sigma^2 I$, σ^2 *known*. Results for this case can be obtained by letting $n \to \infty$. $S/(n+2)$ is replaced by σ^2; i.e., $F = X'X/S$ becomes $(n+2)X'X/\sigma^2$, so that, e.g., (1.2) becomes $(1 - (n+2)cX'X/\sigma^2)X$, $0 < c < 2(p-2)/(n+2)$, or, equivalently, $(1 - aX'X/\sigma^2)X$, $0 < a < 2(p-2)$.

Σ *completely unknown*. Now let Y be a p-dimensional ($p \geq 3$) normal random vector with unknown mean vector η and unknown covariance matrix Σ. Let W be distributed independently of Y according to the Wishart distribution with m degrees of freedom and parameter Σ. The problem is to estimate η when the loss function is

$$L^*(\hat{\eta}; \eta, \Sigma) = (\hat{\eta} - \eta)' \Sigma^{-1} (\hat{\eta} - \eta) .$$

Denote estimators for η by ϕ; denote risks with respect to L^* by R^*. Let J be distributed according to the Poisson distribution with parameter $\eta'\Sigma^{-1}\eta/2$. Let $T = Y'W^{-1}Y$. We shall be interested in estimators of the form $\phi_h(Y, W) = h(T)Y$. Such estimators are invariant estimators with respect to the transformation $Y \to CY$, $W \to CWC'$, where C is a $p \times p$ nonsingular matrix. (Let Y_1, $Y_2, \cdots, Y_n$ be independent and identically distributed according to a p-variate normal distribution with mean vector η and covariance matrix Γ. Then $Y = \bar{Y} \equiv \sum_{i=1}^n Y_i/N$ is p-variate normal with mean vector η and covariance matrix $\Sigma = \Gamma/N$, and $W = \sum_{1=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'/N$ is distributed independently of Y according to a Wishart distribution with parameter Σ and $m = N - 1$ degrees of freedom. If we make the transformation $Y_i \to CY_i$, then the induced transformation of Y and W is $Y \to CY$, $W \to CWC'$.)

LEMMA 4. $R^*(\phi_h; \eta, \Sigma) = E[h^2(\chi_{p+2J}^2/\chi_{m-p+1}^2)\chi_{p+2J}^2] - 4E[h(\chi_{p+2J}^2/\chi_{m-p+1}^2)J] + \eta'\Sigma^{-1}\eta$, *where χ_{p+2J}^2 and χ_{m-p+1}^2 are independent.*

PROOF. The risk is

$$R^*(\phi_h; \eta, \Sigma) = E[(h(Y'W^{-1}Y)Y - \eta)'\Sigma^{-1}(h(Y'W^{-1}Y)Y - \eta)] .$$

Let A be a matrix such that $A\Sigma A' = I$ and make the transformation $Y^* = AY$, giving $W^* = AWA'$. Y^* is normal with mean vector $\eta^* = A\eta$ and covariance matrix I. W^* is distributed independently of Y^* according to the Wishart distribution with m degrees of freedom and parameter I. The risk is $E[\|h(Y^{*\prime}W^{*-1}Y^*)Y^* - \eta^*\|^2]$. It can be shown that the conditional distribution of $Y^{*\prime}W^{*-1}Y^*$ given Y^* is that of $\|Y^*\|^2/S^*$, where S^* is distributed independently of Y^* as χ_{m-p+1}^2. (See Anderson, [1] page 106.) $\|Y^*\|^2$ is distributed

as noncentral chi-square with p degrees of freedom and noncentrality $\eta^{*\prime}\eta^* = \eta'\Sigma^{-1}\eta$, i.e., as x^2_{p+2J}. Thus the risk is $E[||h(Y^{*\prime}Y^*/S^*)Y^* - \eta^*||^2]$, and the result now follows by application of Lemma 1.

REMARK. An invariance argument shows that the risk of $h(T)Y$ depends upon the parameters only through $\eta'\Sigma^{-1}\eta$; thus the risk is a function of $\eta'\Sigma^{-1}\eta$, m and p.

THEOREM 4. $R^*(\varphi_h; \eta, \Sigma) = f(\eta'\Sigma^{-1}\eta, m - p + 1, p)$ *if and only if* $R(\varphi_h; \theta, \sigma^2) = f(\theta'\theta/\sigma^2, n, p)$.

PROOF. (Follows at once by comparing Lemmas 4 and 1.)

COROLLARY. *The estimator $k(F)X$ dominates $h(F)X$ for estimating θ if and only if $k(T)Y$ dominates $h(T)Y$ for estimating η.*

Thus, for example, the fact that $(1 - d/Y'W^{-1}Y)Y$, $0 < d < 2(p - 2)/(m - p + 3)$, the estimator discussed in [2], is improved by replacing $1 - d/Y'W^{-1}Y$ by its positive part follows from the corresponding fact for $(1 - cS/X'X)X$.

Acknowledgment. We wish to thank Mary Ellen Bock, who pointed out an error in the proof of the original version of Theorem 3, our correction of which resulted in the present, more general version of Theorem 3.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.
[2] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.
[3] STEIN, C. (1966). An approach to the recovery of interblock information in balanced incomplete block designs. *Research Papers in Statistics: Festschrift for J. Neyman,* (F. N. David ed.). 351–366. Wiley, New York.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305