ESTIMATES OF LINEAR COMBINATIONS OF THE PARAMETERS IN THE MEAN VECTOR OF A MULTIVARIATE DISTRIBUTION[1]

By Arthur Cohen

Rutgers—The State University

1. Introduction. Let y be an observation on a $p \times 1$ random vector whose distribution is multivariate normal with mean vector θ and known nonsingular covariance matrix Φ. Let φ be a $p \times 1$ vector of constants. Suppose we wish to estimate $\varphi'\theta$ when the loss function is squared error. For the case $p = 1$, it follows from a result due to Karlin [5] that a linear estimate of the form γy is admissible if and only if γ lies in the interval $[0, \varphi]$. In this paper we generalize this result to the case of arbitrary p by proving that $\gamma' y$ is admissible if and only if the $p \times 1$ vector γ lies in or on the ellipsoid

$$(\gamma - \varphi/2)'\Phi(\gamma - \varphi/2) \leqq \varphi'\Phi\varphi/4.$$

In proving this result we will identify linear estimates which are better than the inadmissible linear estimates, thereby adding to the practicality of the result.

The model assumed in this paper is appropriate for the problem of predicting from a regression function. Another result given here is concerned with the problem of including or deleting the pth variate of a pth order regression to be used for prediction. It is shown that the predictor which depends on the outcome of a significance test is inadmissible. That is, the following predictor is inadmissible: If the absolute value of the pth sample regression coefficient exceeds a given constant, predict by the "usual" linear combination of the p sample regression coefficients; otherwise predict by the "usual" linear combination of the first $(p - 1)$ sample regression coefficients. This type of predictor has been used in practice and has been studied by Bancroft and Larson [2], and others. (See Kitigawa [6], where other references are given.) In proving this inadmissibility result we do not identify any predictor which is better than the predictor in question, thereby limiting the practicality of this finding.

For the problem of predicting from a regression function it is interesting to note that the generalization of Karlin's result mentioned above, implies that the predictor, which modifies the "usual" linear predictor by multiplying the pth sample regression coefficient by a small constant, is admissible. This latter type of predictor has been suggested by Tukey [11].

We remark at this point that the decision theory terminology and definitions used in this paper are more or less that of Blackwell and Girshick [3]. Now note that there is no loss in generality for the problem at hand when we consider only a single observation on y. This is so, since for N independent observations, the

Received 19 March 1964; revised 14 August 1964.

[1] This paper is part of the thesis submitted in partial fulfillment of the requirements for the Ph.D. degree, Columbia University, September, 1963.

sample mean is a sufficient statistic, whose distribution is multivariate normal with mean vector θ and covariance matrix Φ/N. Also for the problem of estimation with squared error loss function, it is well known that the nonrandomized estimates form a complete class. Furthermore, for the squared error loss function, the problem of predicting a function of parameters, which represents the expected value of a random variable, is essentially the same as predicting the random variable. This follows since the risk for the latter problem equals a constant plus the risk for the former problem. Hence we will use prediction and estimation interchangeably, always meaning estimation of the function of the parameters.

In this paper we use the term linear estimate meaning a homogeneous linear estimate. The results could be transformed to results for non-homogeneous linear estimates. For such a case, estimates $\gamma' y + k$, for any constant k, are admissible if and only if γ lies in or on the ellipsoid

$$(\gamma - \varphi/2)'\Phi(\gamma - \varphi/2) \leqq \varphi'\Phi\varphi/4,$$

save for the case $\gamma = \varphi$. That is, if $\gamma = \varphi$, $\gamma' y + k$ is admissible if and only if $k = 0$.

In the next section we prove the generalization of Karlin's result. In Section 3 we prove that the predictor which depends on the outcome of a significance test is inadmissible. Finally, in Section 4 we conclude the paper with some generalizations and discussion.

2. All admissible linear estimates. In this section we find all admissible linear estimates. We do this by first finding the admissible estimates for the case when the covariance matrix Φ is the identity matrix of rank p. For Theorems 2.1, 2.2, and 2.3 to follow then, we assume Φ is the identity matrix of rank p. In Theorem 2.1 we show that $\gamma' y$ is inadmissible whenever γ lies outside the sphere

$$(\gamma - \varphi/2)'(\gamma - \varphi/2) \leqq \varphi'\varphi/4. \tag{2.1}$$

As a result of Theorem 2.2, we will see that $\gamma' y$ is an essentially unique Bayes solution for a specified *a priori* distribution whenever γ lies inside the sphere defined in (2.1). Hence such estimates are admissible. In Theorem 2.3, we apply a theorem due to Stein [9] to prove that $\gamma' y$ is admissible when γ lies on the sphere given in (2.1). Finally, by use of a transformation, we obtain the result we are seeking in Theorem 2.4.

To start we prove

LEMMA 2.1. *The sphere given in* (2.1) *is contained in the sphere*

$$\gamma'\gamma \leqq \varphi'\varphi. \tag{2.2}$$

PROOF. If γ lies in or on the sphere given in (2.1), then

$$\gamma'\gamma \leqq \gamma'\varphi. \tag{2.3}$$

The lemma now follows by applying the Cauchy inequality to the right-hand side of (2.3). Next we prove

THEOREM 2.1. *Any estimate $\gamma' y$ where γ lies outside the sphere given in* (2.1) *is inadmissible.*

PROOF. First note that the risk function for any procedure of the form $\gamma' y$ is

$$\rho(\theta, \gamma' y) = \gamma'\gamma + ((\gamma - \varphi)'\theta)^2. \tag{2.4}$$

Now suppose we consider $\tilde{\gamma}' y$ where $\tilde{\gamma}$ lies outside the sphere given in (2.2). By Lemma 2.1 it follows that $\tilde{\gamma}$ also lies outside the sphere given in (2.1). Now for such a $\tilde{\gamma}$, $\varphi' y$ is obviously better than $\tilde{\gamma}' y$ and so $\tilde{\gamma}' y$ is inadmissible.

Now choose $\tilde{\gamma}$ so that $\tilde{\gamma}$ lies outside the sphere in (2.1) but lies inside or on the sphere given in (2.2). Then this $\tilde{\gamma}$ satisfies

$$\tilde{\gamma}'\varphi < \tilde{\gamma}'\tilde{\gamma} \leqq \varphi'\varphi. \tag{2.5}$$

We assert that the procedure $\gamma^{*\prime} y$ is better than $\tilde{\gamma}' y$, where $\gamma^* = c(\tilde{\gamma} - \varphi) + \varphi$ for

$$c = (\varphi'\varphi - \tilde{\gamma}'\varphi)/(\tilde{\gamma} - \varphi)'(\tilde{\gamma} - \varphi). \tag{2.6}$$

For from (2.4)

$$\rho(\theta, \gamma^{*\prime} y) = \gamma^{*\prime}\gamma^* + c^2((\tilde{\gamma} - \varphi)'\theta)^2,$$

and so if we can show

$$0 \leqq c < 1 \tag{i}$$

and

$$\gamma^{*\prime}\gamma^* < \tilde{\gamma}'\tilde{\gamma}, \tag{ii}$$

then our assertion will be correct. That (i) is true follows from (2.5). With regard to (ii), if we write out the expression for $\gamma^{*\prime}\gamma^*$ substituting the right-hand side of (2.6) for c, we find that (ii) reduces to

$$\varphi'\varphi - (\varphi'(\varphi - \tilde{\gamma}))^2/(\tilde{\gamma} - \varphi)'(\tilde{\gamma} - \varphi) < \tilde{\gamma}'\tilde{\gamma}. \tag{2.7}$$

If we transpose, expand, and reduce, we find (2.7) can be written as

$$0 < (\tilde{\gamma}'\tilde{\gamma})^2 + (\varphi'\tilde{\gamma})^2 - 2(\tilde{\gamma}'\tilde{\gamma})(\varphi'\tilde{\gamma}) = (\tilde{\gamma}'\tilde{\gamma} - \varphi'\tilde{\gamma})^2. \tag{2.8}$$

But again from (2.5) it follows that (2.8) is true and so we have verified (ii) and thus have completed the proof of Theorem 2.1. We next prove

LEMMA 2.2. *If the a priori distribution of θ is multivariate normal with mean vector zero and covariance matrix Σ, then the conditional expectation of θ given y is*

$$E(\theta \mid y) = \Sigma(\Sigma + I)^{-1} y.$$

PROOF. Suppose Σ is of rank $r \leqq p$. Then there exists a nonsingular matrix B such that

$$B\Sigma B' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

where I_r is the identity matrix of rank r. Let $V = B\theta$. Then $V = \binom{V^{(1)}}{V^{(2)}}$, where $V^{(1)}$ is an $r \times 1$ vector which is multivariate normal with mean vector zero and covariance matrix I_r, and $V^{(2)}$ is a $(p - r) \times 1$ vector of zeros. Now suppose we partition B into $\binom{B_1}{B_2}$ where B_1 is $r \times p$ and B_2 is $(p - r) \times p$, and we also partition B^{-1} into (C, D), where C is $p \times r$ and D is $p \times (p - r)$. Then $\theta = B^{-1}V = CV^{(1)}$. Now suppose we consider the $(p + r) \times 1$ vector $\binom{y}{V^{(1)}}$. It is easy to verify (see [1], p. 29) that this vector is multivariate normal with mean vector zero and covariance matrix

$$\begin{pmatrix} \Sigma + I & C \\ C' & I_r \end{pmatrix}.$$

This implies that $V^{(1)}$ given y is multivariate normal with mean vector

$$E(V^{(1)} \mid y) = C'(\Sigma + I)^{-1}y.$$

Since $V^{(2)}$ is a zero vector, and $V^{(2)}$ given y is a zero vector, it follows that

$$E(V \mid y) = (C'(\Sigma + I)^{-1}y).$$

Now from an elementary property of conditional expectation we have

$$E(B^{-1}V \mid y) = B^{-1}E(V \mid y)$$

and hence it follows that

$$E(B^{-1}V \mid y) = E(\theta \mid y) = CC'(\Sigma + I)^{-1}y = \Sigma(\Sigma + I)^{-1}y.$$

This proves the lemma for any r, $1 \leqq r \leqq p$. Now we are ready to prove

THEOREM 2.2. *Any estimate $\gamma' y$, where γ lies inside the sphere given in* (2.1) *is an essentially unique Bayes solution for some a priori distribution.*

PROOF. We shall show that if γ lies inside the sphere given in (2.1), then $\gamma' y$ is an essentially unique Bayes solution for an *a priori* distribution which is normal with mean vector zero and covariance matrix Σ, where Σ has rank one. It will also be clear, from the technique of the proof, that $\gamma' y$ is an essentially unique Bayes solution for other *a priori* distributions which are normal with mean vector zero and covariance matrix of rank r, for all r such that, $1 \leqq r \leqq p$.

To prove Theorem 2.2, we first note that from Lemma 2.2, and the fact that the essentially unique Bayes solution is the *a posteriori* expected value of $\varphi'\theta$, that the Bayes solution with respect to an *a priori* distribution which is multivariate normal with mean vector zero and covariance matrix Σ is $\varphi'(\Sigma(\Sigma + I)^{-1}y)$. Therefore, if we are given a γ which lies inside the sphere given in (2.1) and we can find a covariance matrix Σ (that is, a positive semi-definite symmetric matrix) such that

$$\gamma = (\Sigma + I)^{-1}\Sigma\varphi, \tag{2.9}$$

then it will follow that $\gamma' y$ is an essentially unique Bayes solution.

Note that we can write (2.9) as

$$\gamma = \Sigma(\varphi - \gamma). \tag{2.10}$$

Also since any positive semi-definite symmetric matrix Σ is such that

$$\Sigma = Q\Delta Q',$$

where Q is an orthogonal matrix and Δ is a positive semi-definite diagonal matrix, we can write condition (2.10) as

$$\gamma = Q\Delta Q'(\varphi - \gamma), \tag{2.11}$$

and if we multiply (2.11) on the left by Q', we get

$$Q'\gamma = \Delta Q'(\varphi - \gamma). \tag{2.12}$$

Thus, for the given vector γ, we need to show the existence of appropriate matrices Q' and Δ satisfying (2.12). Now notice that if γ lies inside the sphere given in (2.1), this implies that $\gamma'(\varphi - \gamma) > 0$, which in turn implies that there exists an orthogonal transformation, Q^* say, such that

$$(Q^{*\prime}\gamma)' = (x_1, 0, \cdots, 0) \quad \text{and} \quad (Q^{*\prime}(\varphi - \gamma))' = (z_1, z_2, \cdots, z_p)$$

where $x_1z_1 > 0$. Hence if we let $d_1 = x_1/z_1$, $d_i = 0$, $i = 2, 3, \cdots, p$, and define the diagonal elements of Δ to be d_i, $i = 1, 2, \cdots, p$, then Δ and Q^* are existing matrices that satisfy (2.12). Hence, if we set $\Sigma^* = Q^{*\prime}\Delta Q^*$, we have shown the existence of a positive semi-definite matrix satisfying (2.10), thus completing the proof.

Next we proceed to prove the admissibility of $\gamma'y$ for those γ lying on the sphere given in (2.1). We shall use a theorem and remark by Stein [9], which we state as

LEMMA 2.3. *Let B be the σ-algebra of all Borel subsets of the real line U, and C a σ-algebra of subsets of a set V. Let μ be Lebesgue measure on B and ν a probability measure on C. Let f be a nonnegative valued BC measurable function on $U \times V$ such that*

$$\int f(u, v)\, du = 1 \qquad \textit{for all } v,$$

$$\int uf(u, v)\, du = 0 \qquad \textit{for all } v;$$

$$\int d\nu\left(\int u^2 f(u, v)\, du\right)^{\frac{3}{2}} < \infty,$$

where we write du instead of $d\mu(u)$. If we observe (U, V) distributed so that for some unknown ξ, $(U - \xi, V)$ has probability density f with respect to $\mu\nu$, then U is an admissible estimator of ξ with squared error as loss. Furthermore, if $g(u, v)$ is any BC measurable function on $U \times V$ such that its risk is less than or equal to the risk of the estimator U, for all ξ, then $g(u, v) = u$ almost everywhere $(\mu\nu)$.

Now we prove

THEOREM 2.3. *If γ lies on the sphere given in (2.1), then $\gamma'y$ is admissible.*

PROOF. Suppose $\gamma'y$ is not admissible. Then there exists some estimate, say $g(y)$, which is better. The risk function for $g(y)$ is

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (g(y) - \varphi'\theta)^2 \\
&\quad\cdot(2\pi)^{-p/2}\exp\left[-(y-\theta)'(y-\theta)/2\right] dy_1\, dy_2 \cdots dy_p = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \\
&\quad\cdot(g(y) - \gamma'\theta)^2(2\pi)^{-p/2}\exp\left[-(y-\theta)'(y-\theta)/2\right] dy_1\, dy_2 \cdots dy_p \\
&\quad + 2(\gamma - \varphi)'\theta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (g(y) - \gamma'\theta) \\
&\quad\cdot(2\pi)^{-p/2}\exp\left[-(y-\theta)'(y-\theta)/2\right] dy_1\, dy_2 \cdots dy_p + ((\gamma - \varphi)'\theta)^2.
\end{aligned} \tag{2.13}$$

Note that since γ lies on the sphere in (2.1) it follows that $\gamma'(\gamma - \varphi) = 0$ and hence we may define an orthogonal $p \times p$ matrix Γ whose first row is $\gamma'/(\gamma'\gamma)^{\frac{1}{2}}$ and whose second row is $(\gamma - \varphi)'/[(\gamma - \varphi)'(\gamma - \varphi)]^{\frac{1}{2}}$. Now let $z = \Gamma y$, and $\omega = \Gamma\theta$, so that the right-hand side of (2.13) becomes

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[g(\Gamma'z) - \omega_1(\gamma'\gamma)^{\frac{1}{2}}]^2 \\ &\quad\cdot(2\pi)^{-p/2}\exp[-(z-\omega)'(z-\omega)/2]\,dz_1\,dz_2\cdots dz_p \\ &+ 2\omega_2[(\gamma-\varphi)'(\gamma-\varphi)]^{\frac{1}{2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[g(\Gamma'z) - \omega_1(\gamma'\gamma)^{\frac{1}{2}}] \\ &\quad\cdot(2\pi)^{-p/2}\exp[-(z-\omega)'(z-\omega)/2]\,dz_1\,dz_2\cdots dz_p \\ &+ \omega_2^2(\gamma-\varphi)'(\gamma-\varphi). \end{aligned} \tag{2.14}$$

If $g(\Gamma'z)$ is better than $\gamma'\Gamma'z$ its risk must always be less than or equal to the risk of $\gamma'\Gamma'z$ for all θ or equivalently, for all ω. In particular, since the coordinates $\omega_1, \omega_2, \cdots, \omega_p$ are orthogonal, the risk for $g(\Gamma'z)$ for $\omega_2 = 0, \omega_3 = 0, \cdots, \omega_p = 0$, must be less than or equal to $\gamma'\gamma$, which is the risk for $\gamma'y$ at $\omega_2 = \omega_3 = \cdots = \omega_p = 0$. That is, from (2.14) we must have

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[g(\Gamma'z) - \omega_1(\gamma'\gamma)^{\frac{1}{2}}]^2 \\ &\quad\cdot(2\pi)^{-\frac{1}{2}}\exp[-(z_1-\omega_1)^2/2]\,dz_1\prod_{i=2}^{p}(2\pi)^{-\frac{1}{2}}\exp[-z_i^2/2]\,dz_i \leqq \gamma'\gamma, \end{aligned}$$

which may be written as

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[g(\Gamma'z)/(\gamma'\gamma)^{\frac{1}{2}} - \omega_1]^2 \\ &\quad\cdot(2\pi)^{-\frac{1}{2}}\exp[-(z_1-\omega_1)^2/2]\,dz_1\prod_{i=2}^{p}(2\pi)^{-\frac{1}{2}}\exp[-z_i^2/2]\,dz_i \leqq 1. \end{aligned} \tag{2.15}$$

Now if we apply Lemma 2.3 with $f(u, v) = \exp[-(u - Eu)^2/2]/(2\pi)^{\frac{1}{2}}$ and with ν equal to the probability measure corresponding to the $(p - 1)$ dimensional normal distribution with mean vector zero and covariance matrix the identity matrix of rank $(p - 1)$, we find that the only function (except for sets of p-dimensional Lebesgue measure zero) $g(\Gamma'z)$ satisfying (2.15) is $g(\Gamma'z) = (\gamma'\gamma)^{\frac{1}{2}}z_1 = \gamma'y$. Hence we may conclude that $g(y)$ cannot be better than $\gamma'y$ and therefore $\gamma'y$ is admissible. This completes the proof of Theorem 2.3.

We conclude Section 2 by considering the case where y is multivariate normal with mean vector θ and with known nonsingular covariance matrix Φ. We prove

THEOREM 2.4. *The estimate $c'y$ is admissible if the $p \times 1$ vector of constants c lies in or on the ellipsoid.*

$$(c - \varphi/2)'\Phi(c - \varphi/2) \leqq \varphi'\Phi\varphi/4; \tag{2.16}$$

otherwise $c'y$ is inadmissible.

PROOF. Let A be the nonsingular $p \times p$ matrix such that $A\Phi A' = I$. Let $z = Ay$ and $\omega = A\theta$. Than z is multivariate normal with mean vector ω and covariance matrix I. Now let $\varphi^* = (A^{-1})'\varphi$, so that $\varphi^{*\prime}\omega = \varphi'A^{-1}\omega = \varphi'\theta$. Then it follows from Theorems 2.1, 2.2, and 2.3 that $\gamma'z = (\gamma'A)y$ is an admissible estimate of $\varphi'\theta$ provided the vector of constants γ lies in or on the sphere

$$(\gamma - \varphi^*/2)'(\gamma - \varphi^*/2) \leqq \varphi^{*\prime}\varphi^*/4; \tag{2.17}$$

otherwise $\gamma' z$ is inadmissible. Note that we may rewrite (2.17) as

$$(2.18) \qquad (\gamma - (A^{-1})'\varphi/2)'(\gamma - (A^{-1})'\varphi/2) \leqq \varphi'\Phi\varphi/4.$$

If we set $c = A'\gamma$, we see from (2.18) that $c'y$ is an admissible estimate for $\varphi'\theta$ provided the vector c lies in or on the ellipsoid in (2.16), and $c'y$ is inadmissible otherwise. This completes Theorem 2.4.

3. Inadmissibility of predictor that depends on a significance test. We now show that a predictor which includes or excludes the pth variate of a pth order regression, depending on the outcome of a significance test is inadmissible. More formally, let the $p \times 1$ vector y be normally distributed with mean vector θ and known nonsingular covariance matrix Φ. We require that the covariances of (y_i, y_p) are zero for all $i = 1, 2, \cdots, p - 1$. (We may think of y as the vector of least squares estimates of θ, where θ is the vector of parameters in a general linear hypothesis model of full rank. The restriction on the covariances of (y_i, y_p) is not severe since the regression variates could be transformed so that the covariance matrix has this property.) Suppose then that we wish to predict $\varphi'\theta$ for some vector of constants φ and we regard the loss function to be squared error. Let us consider the predictor $t(y)$ defined as follows:

$$(3.1) \qquad \begin{aligned} t(y) &= \textstyle\sum_{j=1}^{p} y_j\varphi_j \qquad \text{if } |y_p| \geqq C \\ &= \textstyle\sum_{j=1}^{p-1} y_j\varphi_j \qquad \text{if } |y_p| < C \end{aligned}$$

for C a positive constant. When $p = 1$, $t(y) = 0$ if $|y_1| < C$. We now prove

THEOREM 3.1. *The predictor $t(y)$ is inadmissible.*

PROOF. For the case $p = 1$, it follows from a result by Sacks [7] (See in particular the last paragraph on p. 767), that a necessary condition for the admissibility of a predictor is that it be an analytic function of y_1, or must be equivalent to an analytic function; that is, it may differ from some analytic function on a set of Lebesgue measure zero. But note that $t(y)$ is not equivalent to any analytic function. For, no matter how we alter $t(y)$ on a set of measure zero, we will always have

$$\underline{\lim}_{y \to C}\, t(y) \leqq 0,$$

and when φ_1 is positive,

$$\overline{\lim}_{y \to C}\, t(y) = C\varphi_1 > 0.$$

When φ_1 is negative, we could give a similar argument. Thus $t(y)$ cannot be an admissible predictor.

For the case of arbitrary p, we can reduce the problem to a one-dimensional problem and again apply results of Sacks [7]. That is, suppose we consider the problem of predicting $\varphi'\theta$ by predictors of the form

$$\delta(y) = \textstyle\sum_{j=1}^{p-1} y_j\varphi_j + f(y_p),$$

where $f(y_p)$ is some function of y_p. Clearly, if $t(y)$ is inadmissible for this reduced problem, it will be inadmissible for the original problem. For the reduced problem, if we consider the expected risk for a procedure, for some arbitrary *a priori* distribution $\xi(\theta)$, it follows that any Bayes solution must be of the form

$$\sum_{j=1}^{p-1} y_j\varphi_j + \varphi_p \int_{-\infty}^{\infty} \theta_p \exp\left[-(y_p - \theta_p)^2/2\sigma_{pp}\right] d\xi(\theta_p) \Big/ \int_{-\infty}^{\infty} \exp\left[-(y_p - \theta_p)^2/2\sigma_{pp}\right] d\xi(\theta_p),$$

where $\xi(\theta_p)$ represents the marginal distribution of θ_p.

Now, by virtue of the proof of Sacks' results (See [7], Section 2, pp. 754–765) any procedure that is the limit (regular sense) of a sequence of Bayes procedures must be of the form

$$\sum_{j=1}^{p-1} y_j\varphi_j + f(y_p), \tag{3.2}$$

where $f(y_p)$ is an analytic function of y_p. Also, as mentioned in Sacks [7], only procedures which are the limits of Bayes procedures can be admissible. Since $t(y)$ is not of the form (3.2) we conclude that $t(y)$ cannot be admissible. This completes the proof of Theorem 3.1.

4. Generalizations and discussion. The first generalization we recognize is concerned with the result of Section 2. If we relax the assumption that y is multivariate normal, assuming only that the covariance matrix of y exists, and if we restrict ourselves to linear estimates, then Theorem 2.4 is still true. This is so, since we note that the risk function for linear estimates depends only on first and second moments. Hence, the proof of Theorem 2.1 never requires the normality assumption. Furthermore, if $c'y$ were inadmissible for some c lying in or on the ellipsoid in (2.16), then from Theorem 2.1 there would have to exist a c^* lying in or on the ellipsoid such that $c^{*\prime}y$ would be better than $c'y$. But if this were true, it would also be true in the case where y was normally distributed, contradicting the result of Theorem 2.4.

The next generalization, which is somewhat obvious, is concerned with the problem of estimating the individual θ_i, $i = 1, 2, \cdots, p$, where y is multivariate normal with mean vector θ, covariance matrix the identity and when the loss function is the sum of the squared errors. That is, the loss function is

$$W(\theta, \delta) = \sum_{i=1}^{p} (\delta_i(y) - \theta_i)^2,$$

where $\delta' = (\delta_1(y), \cdots, \delta_p(y))$, and $\delta_i(y)$ is an estimate of θ_i. If we use the results of Karlin [5] and Stein [8] it is easy to show that the estimates of the form $G'(y) = (\gamma_1 y_1, \gamma_2 y_2, \cdots, \gamma_p y_p)$ are admissible if and only if $0 \leqq \gamma_j \leqq 1$, for all $j = 1, 2, \cdots, p$ and $\gamma_j = 1$ for at most two of the indices, $j = 1, 2, \cdots, p$. A question that now arises for this case is, can we find all the admissible estimates of the form $\Gamma(y)$, where

$$\Gamma'(y) = (\gamma_{11}y_1 + \gamma_{12}y_2 + \cdots + \gamma_{1p}y_p,\ \gamma_{21}y_1 + \gamma_{22}y_2 + \cdots + \gamma_{2p}y_p, \cdots, \gamma_{p1}y_1 + \gamma_{p2}y_2 + \cdots \gamma_{pp}y_p)?$$

Some final generalizations are concerned with the following problem: Let z_i, $i = 1, 2, \cdots, n$, be independent observations on a random vector z, where

$z' = (y, x)$, is bivariate normal with mean vector zero and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \sigma_{xx} \end{pmatrix}.$$

Suppose we assume that σ_{xx} and $\sigma^2 = \sigma_{yy} - \sigma_{yx}^2/\sigma_{xx}$ are known (and without loss of generality these variances are set equal to 1). If we denote σ_{xy}/σ_{xx} by β and consider the problem of estimating β when the loss function is squared error, it is possible to prove

THEOREM 4.1. *The estimate*

$$\Lambda(z) = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2 + \lambda,$$

for $0 < \lambda < \infty$, *and for* $\lambda = 0$ *and* $n \geqq 4$, *is admissible; while the estimate*

$$t(z) = \hat{\beta} \quad \text{if } |\hat{\beta}| \geqq C$$
$$= 0 \quad \text{if } |\hat{\beta}| < C,$$

where $\hat{\beta} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2$, *C is a positive constant, is inadmissible.*

The proof follows by showing $\Lambda(z)$ is an essentially unique Bayes solution whenever $0 < \lambda < \infty$, and when $\lambda = 0$, the result has been proved for $n \geqq 4$ by Stein [10]. The proof of the inadmissibility of $t(z)$ uses essentially the same arguments given to prove Theorem 3.1.

In the introduction we made a remark concerning the problem of deleting or including the pth variate of a pth order regression to be used for prediction. We said that whereas the predictor $t(y)$ given in (3.1) is inadmissible for the squared error loss function, the predictor which modifies the "usual" linear predictor by multiplying the pth sample regression coefficient by a small constant is admissible. One may think of this perhaps as being a compromise between leaving the pth sample regression coefficient in or out. It is also interesting to point out that in the author's dissertation [4] a formulation is given for a problem where one decides if the regression is of the pth order or lower, and then predicts. For the formulation given there a predictor of the type $t(y)$ given in (3.1) is found to be admissible.

5. Acknowledgment. I am deeply grateful to Professor T. W. Anderson for his guidance throughout the course of this work. His suggestions, comments, and criticisms were of great value to me. This work was done while under a grant from the Biostatistics Training Program of Columbia University, which is partially supported by the National Institutes of Health. I would like to take this opportunity to thank Professor J. W. Fertig, Chairman of the Biostatistics Department, for his continual encouragement.

REFERENCES

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

[2] BANCROFT, T. A. and LARSON, H. J. (1963). Biases in prediction by regression for certain incompletely specified models. *Biometrika* **50** 391–402.

[3] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[4] COHEN, ARTHUR (1963). Some questions of predicting from a regression function and a hybrid problem on the exponential family. Ph.D. thesis, Columbia Univ.

[5] KARLIN, SAMUEL (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.

[6] KITIGAWA, TOSIO (1963). Estimation after preliminary tests of significance. *Univ. California Publ. Statist.* **3** 147–186.

[7] SACKS, JEROME (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* **34** 751–768.

[8] STEIN, CHARLES (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

[9] STEIN, CHARLES (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Math. Statist.* **30** 970–979.

[10] STEIN, CHARLES (1960). Multiple regression. *Contributions to Probability and Statistics —Essays in Honor of Harold Hotelling.* 424–443. Stanford Univ. Press.

[11] TUKEY, JOHN W. (1960). Where do we go from here? *J. Amer. Statist. Assn.* **55** 80–93.