MULTIPLE REGRESSION AND ESTIMATION OF THE MEAN
OF A MULTIVARIATE NORMAL DISTRIBUTION

BY
ALVIN J. BARANCHIK

TECHNICAL REPORT NO. 51
MAY 4, 1964

PREPARED UNDER CONTRACT Nnonr-225(72)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MULTIPLE REGRESSION AND ESTIMATION OF THE MEAN
OF A MULTIVARIATE NORMAL DISTRIBUTION

by
Alvin J. Baranchik

TECHNICAL REPORT NO. 51
May 4, 1964

PREPARED UNDER CONTRACT Nonr-225(72)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SECTION 1
INTRODUCTION

This paper is concerned with two problems: estimation of the mean vector of a multivariate normal random variable, and estimating the regression function in the case of one predictand and at least three predictors. The first of these problems takes up section 2 and the latter, section 3.

In 1956 Stein [4] proved that the sample mean, long the customary estimator, was inadmissible as an estimator of the mean of a multivariate normal random variable, when the number of variables was at least 3. James and Stein [2] later gave an explicit estimator for the mean with everywhere lower risk than that of the usual estimator, but their estimator is still inadmissible. In section 2 of this paper a family of minimax estimators is given, at least one of which is admissible. The relationship of these estimators to those given by James and Stein [2] is also studied, and it is tentatively concluded that the newly computed estimators would not be of great practical value. Section 2 also includes a study of the component-wise risk of the estimators proposed in [2], and a theorem which yields a simple improvement over those estimators.

Section 3 gives estimators which have strictly lower risks than the corresponding maximum likelihood estimators for the following three problems:

(i) estimation of the regression function,

(ii) estimation of the regression coefficients, and
(iii) estimation of the mean of the predictand when the means of the predictor variables are known.

In each of these problems it is assumed that there are at least three predictor variables. A summary of previously known results for these three problems is given at the beginning of section 3.1. Section 3 concludes with some elementary modifications of the estimators computed earlier in that section. It is believed that these modified estimators will frequently be of practical use.
SECTION 2

ESTIMATING THE MEAN OF A MULTIVARIATE NORMAL RANDOM VARIABLE

2.1 A family of estimators.

Let $X$ be a $p \times 1$ random vector distributed as a multivariate normal with mean vector $\theta$ and known covariance matrix $I_p$, the $p \times p$ identity matrix, viz., $E(x-\theta)(x-\theta)' = I_p$. Traditionally, the mean vector has been estimated by

\begin{equation}
\hat{\theta} = \bar{x},
\end{equation}

the maximum likelihood estimator. With the sum of squared errors loss function

\begin{equation}
L(\hat{\theta}; \theta) = \sum_{i=1}^{p} (\hat{\theta}_i, \theta) = \|\hat{\theta} - \theta\|^2,
\end{equation}

where throughout this paper we shall understand $\theta_i$ to be the $i^{th}$ coordinate of the column vector $\theta$ and $\|\theta\|^2$ to be $\Sigma \theta_i^2$, the estimator $\hat{\theta} = \bar{x}$ is of constant risk (expected loss) $p$ and is minimax, that is,

$$\sup_{\theta} \rho(\hat{\theta}, \theta) = \sup_{\theta} \mathbb{E}L(\hat{\theta}, \theta) \leq \sup_{\theta} \mathbb{E}L(\varphi(x), \theta) = \sup_{\theta} \rho(\varphi(x), \theta)$$

for any function $\varphi(x)[2]$.

When $p \leq 2$ (2.1.1) also has the admissibility property, e.g., there does not exist an estimator $\varphi(x)$ with the property

\begin{equation}
\rho(\varphi(x), \theta) < \rho(\hat{\theta}, \theta) \quad \text{with} \quad \rho(\varphi(x), \theta_0) < \rho(\hat{\theta}, \theta_0)
\end{equation}

for some $\theta_0[2]$. Stein [4] showed that, for $p \geq 3$, (2.1.1) is
inadmissible, and James and Stein [2] gave the estimator

\[(2.1.4) \quad \varphi(x) = \left(1 - \frac{p-2}{\|x\|^2}\right)x, \quad \|x\|^2 = \sum_{i=1}^{p} x_i^2,\]

which has the property (2.1.3) and, thus, has everywhere lower risk than \( \hat{\theta} \).

In this subsection we shall derive a family of formal Bayes estimators of \( \theta \), that is, instead of limiting ourselves to prior measures which are probability measures we shall treat certain prior measures on the space \( \{\theta: \, -\infty < \theta_i < \infty, \, i=1,2,\ldots,p\} \) as if they were probability measures and formally compute Bayes estimators with respect to such prior measures.

One such formal measure would be that which distributes \( \theta \) uniformly over \( p \)-dimensional space. This is equivalent to distributing \( \|\theta\|^p \) uniformly over the positive real line, and then selecting a point on the \( p \)-dimensional sphere according to a uniform distribution. This leads (as may be seen as a special case of the computation which follows later in this subsection) to the usual estimator of \( \theta \) (2.1.1). In this case the prior density has differential element \( \prod_{i=1}^{p} d\theta_i \). Analogously, if one were to distribute \( \|\theta\|^{2+\varepsilon} \) uniformly (instead of \( \|\theta\|^p \)) the differential element would be

\[(2.1.5) \quad \|\theta\|^{2+\varepsilon} \prod_{i=1}^{p} d\theta_i.\]

Treating \( \|\theta\|^{2-\varepsilon} \) as a prior probability measure and writing \( f(x) \) for the factor with coordinates \( \frac{df(x)}{dx_i} \), the formal Bayes estimator of \( \theta \) with respect to \( \|\theta\|^{2-\varepsilon} \) is
\[
\varphi_\varepsilon(x) = \frac{f_{\theta^*} - \frac{1}{2} \|x - \theta\|^2}{\int e^{-\frac{1}{2} \|x - \theta\|^2} \|\theta\|^{2-p+\varepsilon} \mathbb{I} \, d\theta} \\
= \nabla \log \left\{ \frac{1}{2} \|x\|^2 \mathbb{E} \chi^2_p(\|x\|^2) \right\}
\]

where $\chi^2_p(\frac{1}{2} \|x\|^2)$ is a chi-square random variable with non-centrality parameter $\frac{1}{2} \|x\|^2$ and $p$ degrees of freedom. Thus

\[
\varphi_\varepsilon(x) = \nabla \log \sum_0^\infty \frac{\left(\frac{1}{2} \|x\|^2\right)^k}{k!} \mathbb{E} \chi^2_p(2k) \int_0^\infty \frac{e^{-\frac{u^2}{2}} - \frac{u}{2}}{\Gamma\left(\frac{p+2k}{2}\right)} (\frac{u}{2})^k \, du \\
= \nabla \log \sum_0^\infty \frac{\left(\frac{1}{2} \|x\|^2\right)^k \Gamma(k+1+\frac{\varepsilon}{2})}{\Gamma\left(\frac{p+2k}{2}\right) k!} \\
= \sum_0^\infty \frac{\Gamma\left(k+1+\frac{\varepsilon}{2}\right)}{\Gamma\left(\frac{p+2k+2}{2}\right) (k+1) \cdot (\frac{1}{2} \|x\|^2)} \cdot x
\]
\begin{align*}
\Omega_{k+1} = \frac{\Gamma(k+2+ \frac{\varepsilon}{2})}{k! \Gamma(\frac{p+2k+2}{2})} \left(\frac{1}{2} \|x\|^2\right)^k \\
\Omega_{k+2} = \frac{\Gamma(k+1+ \frac{\varepsilon}{2})}{k! \Gamma(\frac{p+2k}{2})} \left(\frac{1}{2} \|x\|^2\right)^k
\end{align*}

Setting \( \varepsilon = p-2 \) we, of course, get \( \varphi_{p-2}(x) = x = \hat{\theta} \).

Of particular interest is \( \varphi_o(x) \) which is, for \( p \geq 3 \), an admissible \( 1/ \) estimator of \( \theta \). In the next subsection we shall show that \( \varphi_o(x) \), as well as \( \varphi_\varepsilon(x) \) for other values of \( \varepsilon \), is minimax.

2.2 \textbf{Proof of the minimax property for} \( \varphi_\varepsilon(X) \) \textit{when} \( 0 \leq \varepsilon \leq p-2 \).

In this subsection we shall first prove a lemma which gives a sufficient condition for an estimator of \( \theta \), the mean of a multivariate normal random variable with known covariance matrix \( I_p \), to be minimax. We will then use this lemma to show that the estimators \( \varphi_\varepsilon(X) \), \( 0 \leq \varepsilon \leq p-2 \), are each minimax.

The following lemma generalizes the result that, if \( \varphi_1 \) and \( \varphi_2 \) are minimax estimators, then \( r \varphi_1 + (1-r)\varphi_2 \), where \( r \) is a constant between 0 and 1, is also a minimax estimator. This result is a direct consequence of the convexity of the loss function (2.1.2).

---

1/ C. Stein in private communication.
Lemma. If an estimator may be written in the form

\begin{equation}
(2.2.1) \quad \varphi(x) = h\left(\frac{1}{2} \|x\|^2\right)x = r\left(\frac{1}{2} \|x\|^2\right)(1- \frac{p-2}{\|x\|^2})x + (1 \cdot r\left(\frac{1}{2} \|x\|^2\right))x
\end{equation}

\begin{equation}
= (1 - r\left(\frac{1}{2} \|x\|^2\right)) \frac{p-2}{\|x\|^2}x
\end{equation}

then \( \varphi(x) \) is minimax if

\begin{equation}
(2.2.2) \quad r\left(\frac{1}{2} \|x\|^2\right)
\end{equation}

is a non-decreasing function of \( \|x\| \) and

\begin{equation}
(2.2.3) \quad 0 \leq r\left(\frac{1}{2} \|x\|^2\right) \leq 1.
\end{equation}

Proof. We note in passing that the Lemma is trivial if \( r(\|x\|) = r_0 \), a constant between 0 and 1, since \( X \) and \( (1- \frac{p-2}{\|x\|^2})x \) are both minimax [2]. In what follows we assume \( r(\|x\|) \) is measurable.

To show that \( \varphi(x) \) is minimax we must show the following is not positive for any \( \theta \): Let \( u = \frac{1}{2} \|x\|^2 \),

\begin{equation}
(2.2.4) \quad \mathbb{E}[h\left(\frac{1}{2} \|x\|^2\right)x-\theta]^2 - \mathbb{E}[X-\theta]^2
\end{equation}

\begin{equation}
= 2\mathbb{E}u^2 - 2\theta' \mathbb{E} h(u)X + \|\theta\|^2 - p .
\end{equation}

Now

\begin{equation}
(2.2.5) \quad 2\mathbb{E}u^2 = \frac{1}{2} \|\theta\|^2 \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2} \|\theta\|^2\right)^k}{k!} \mathbb{E} \chi^2_p \mathbb{E} \chi^2_p \chi^2_{p+2k}
\end{equation}

To compute \(-2\theta' \mathbb{E}h(u)X\) we make an orthogonal transformation such that \( \theta' = (\|\theta\|,0,0,\ldots,0)' \). Then
\[
E h(u)X = \frac{1}{(2\pi)^{n/2}} \int \cdots \int h\left(\frac{1}{2} \Sigma y_1^2\right) y_1 e^{-\frac{1}{2} \left(\Sigma y_1^2 - 2\|\theta\|y_1 + \|\theta\|^2\right)} \prod dy_1
\]
\[
= \frac{e^{-\frac{1}{2} \|\theta\|^2}}{(2\pi)^{n/2}} \frac{d}{d\|\theta\|} \left[ \int \cdots \int h\left(\frac{1}{2} \Sigma y_1^2\right) e^{-\frac{1}{2} \left(\Sigma y_1^2 - 2\|\theta\|y_1 + \|\theta\|^2\right)} \prod dy_1 \right]
\]
\[
= \frac{e^{-\frac{1}{2} \|\theta\|^2}}{(2\pi)^{n/2}} \frac{d}{d\|\theta\|} \left[ e^{\frac{1}{2} \|\theta\|^2} \int \cdots \int h\left(\frac{1}{2} \Sigma y_1^2\right) e^{-\frac{1}{2} \left(\Sigma y_1^2 - 2\|\theta\|y_1 + \|\theta\|^2\right)} \prod dy_1 \right]
\]
\[
= e^{-\frac{1}{2} \|\theta\|^2} \frac{d}{d\|\theta\|} e^{\frac{1}{2} \|\theta\|^2} E h\left(\frac{1}{2} \chi_{p+2k}^2\right).
\]

Expanding \(E h\left(\frac{1}{2} \chi_{p+2k}^2\right)\) in a Poisson series we have
\[
E h(u)X = e^{-\frac{1}{2} \|\theta\|^2} \frac{d}{d\|\theta\|} e^{\frac{1}{2} \|\theta\|^2} \sum_{k=0}^{\infty} \frac{\|\theta\|^2}{2^k k!} E h\left(\frac{1}{2} \chi_{p+2k}^2\right)
\]
\[
= e^{-\frac{1}{2} \|\theta\|^2} \sum_{k=0}^{\infty} \theta^k \left(\frac{\|\theta\|^2}{2}\right)^k E h\left(\frac{1}{2} \chi_{p+2k}^2\right) / k!
\]
\[
= \frac{2}{\|\theta\|} \sum_{k=0}^{\infty} e^{-\|\theta\|^2} \left(\frac{\|\theta\|^2}{2}\right)^k E h\left(\frac{1}{2} \chi_{p+k}^2\right) / k!
\]

Combining this value of \(E h(u)X\) with (2.2.5) the difference between the risks of \(\varphi(X)\) and \(X\) (2.2.4) becomes
\[ (2.2.6) \quad e^{-\frac{1}{2} \| \theta \|^2} \sum \frac{(\frac{1}{2} \| \theta \|^2)^k}{k!} \left[ E \chi^2_{p+2k} h^2(\frac{1}{2} \chi^2_{p+2k}) - 4kEh(\frac{1}{2} \chi^2_{p+2k}) \right] \]

\[ + \| \theta \|^2 - p \]

\[ = e^{-\frac{1}{2} \| \theta \|^2} \sum \frac{(\frac{1}{2} \| \theta \|^2)^k}{k!} \left[ E \chi^2_{p+2k} h^2(\frac{1}{2} \chi^2_{p+2k}) - 4kEh(\frac{1}{2} \chi^2_{p+2k}) - p + 2k \right] \]

since \[ e^{-\frac{1}{2} \| \theta \|^2} \sum \frac{(\frac{1}{2} \| \theta \|^2)^k}{k!} \cdot 2k = \| \theta \|^2. \] Our lemma will thus be proved if we can show that

\[ (2.2.7) \quad E\left[ \chi^2_{p+2k} h^2(\frac{1}{2} \chi^2_{p+2k}) - 4kEh(\frac{1}{2} \chi^2_{p+2k}) - k + 2k \right] \leq 0 \]

for each \( k = 0, 1, 2, \ldots \). By the definition of \( h(u) \), (2.2.7) equals

\[ E\left[ \chi^2_{p+2k} h^2(\frac{1}{2} \chi^2_{p+2k}) - p - 2k + 4kr(\frac{1}{2} \chi^2_{p+2k}) \frac{p-2}{\chi^2_{p+2k}} \right] \]

\[ = E\left[ \chi^2_{p+2k} \left( 1 - r(\frac{1}{2} \chi^2_{p+2k}) \right) \frac{p-2}{\chi^2_{p+2k}} - p - 2k + 4kr(\frac{1}{2} \chi^2_{p+2k}) \frac{p-2}{\chi^2_{p+2k}} \right] \]

\[ = E\left[ -2r(\frac{1}{2} \chi^2_{p+2k})(p-2) + \frac{r^2(1/2 \chi^2_{p+2k})}{\chi^2_{p+2k}} + 4kr(1/2 \chi^2_{p+2k}) \frac{p-2}{\chi^2_{p+2k}} \right] \]

because \( E \chi^2_{p+2k} = p+2k \). Since \( r(\frac{1}{2} \chi^2_{p+2k}) = (1-h(\frac{1}{2} \chi^2_{p+2k})) \frac{\chi^2_{p+2k}}{p-2} \) it follows that (2.2.7) equals
\[ E \left[ -2r \left( \frac{1}{2} \chi^2_{p+2k} \right)(p-2) + r \left( \frac{1}{2} \chi^2_{p+2k} \right)(p-2)(1-h(\frac{1}{2} \chi^2_{p+2k})) + 4kr \left( \frac{1}{2} \chi^2_{p+2k} \right) \frac{p-2}{\chi^2_{p+2k}} \right] \]

\[ = E \left[ \left( \frac{4k}{\chi^2_{p+2k}} - 1 - h(\frac{1}{2} \chi^2_{p+2k}) \right)(p-2) \frac{1}{\chi^2_{p+2k}} \right] \]

\[ \leq E \left[ \frac{4k+p-2}{\chi^2_{p+2k}} - 2 \right](p-2) \frac{1}{\chi^2_{p+2k}} \]

since (2.2.3) implies \( h(\frac{1}{2} \chi^2_{p+2k}) \geq 1 - \frac{p-2}{\chi^2_{p+2k}} \). By (2.2.2), \( r(\frac{1}{2} \chi^2_{p+2k}) \) and \( \frac{4k+p-2}{\chi^2_{p+2k}} \) (a decreasing function of \( \chi^2_{p+2k} \)) are not positively correlated. Moreover, \( E \left[ \frac{4k+p-2}{\chi^2_{p+2k}} - 2 \right] = \frac{p-2}{2k+p-2} < 0 \). Therefore, (2.2.7) is not positive for any value of \( k \). Q.E.D.

We are now prepared to prove the theorem:

**Theorem.** \( \varphi_\varepsilon(X) \) (2.1.6) is a minimax estimator of the mean of a multivariate normal with known covariance matrix \( I_p \) if \( 0 \leq \varepsilon \leq p-2 \).

**Proof.** By the lemma it suffices to show that \( r(\frac{1}{2} \|X\|^2) = \|X\|^2 \)

\( (1-h(\frac{1}{2} \|X\|^2)) / (p-2) \) is non-decreasing as a function of \( \|X\|^2 \) and that \( v(\|X\|^2) = \|X\|^2(1-h(\frac{1}{2} \|X\|^2)) \leq p-2 \). Using (2.1.6) we will first prove \( v(\|X\|) \leq p-2 \).
\[ v(|x|) = |x|^2 \left( 1 - \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p+2}{2} \right)} / \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p}{2} \right)} \right) \]

\[ = |x|^2 \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p}{2} \right)} \cdot \frac{\Gamma(j+2+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p+2}{2} \right)} \cdot \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p}{2} \right)} \]

\[ = (p-2-\epsilon) \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^{j+1}}{j! \Gamma \left( \frac{2j+p+2}{2} \right)} / \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p}{2} \right)} \]

since \( \frac{\Gamma(2j+p)}{\Gamma(2j+p+2)} \cdot (2j+p) = \frac{\Gamma(2j+p+2)}{2} \).

Therefore, letting \( j=k-1 \),

\[ v(|x|) = (p-2-\epsilon) \frac{\Gamma(k+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^k}{(k-1)! \Gamma \left( \frac{2k+p}{2} \right)} / \sum_{j=0}^{\infty} \frac{\Gamma(j+1+\frac{\epsilon}{2}) \left( \frac{1}{2} |x|^2 \right)^j}{j! \Gamma \left( \frac{2j+p}{2} \right)} \]

But the series in the numerator has smaller coefficients of \( \left( \frac{1}{2} |x|^2 \right)^j \), \( j=0,1, \ldots \), than has the series in the denominator. Therefore,

\[ v(|x|) \leq p-2-\epsilon \leq p-2, \quad \text{whenever } \epsilon \geq 0. \]
For the second part of the proof we must show \( v(\|x\|) = (p-2)r(\frac{1}{2}\|x\|^2) \) is an increasing function of \( \|x\| \). This will be so if \( \frac{dv(\|x\|)}{d\|x\|} \geq 0 \).

The derivative of \( v(\|x\|) \) will be non-negative whenever

\[
\sum_{j=0}^{\infty} \frac{\Gamma[j+\frac{p}{2}](\frac{1}{2}\|x\|^2)^j}{j! \Gamma(\frac{2j+p}{2})} \geq 0
\]

is positive. Clearly the coefficient of \( (\frac{1}{2}\|x\|^2)^0 \) is positive. We now look at the coefficient of \( (\frac{1}{2}\|x\|^2)^{l+1} \), \( l=0,1,2,\ldots \). Fixing \( j+k=l+1 \) we set \( j=i \) and \( k=l-i+1 \) and then \( j=l-i+1, k=1 \). Adding these two contributing terms to the coefficient of \( (\frac{1}{2}\|x\|^2)^{l+1} \) we obtain
\[
\begin{align*}
\frac{\Gamma(i+1+\frac{\ell}{2})\Gamma(\ell-1+1+\frac{\ell}{2})(\ell-1+1)}{i!\Gamma\left(\frac{2\ell+p}{2}\right)(\ell-1)\Gamma\left(\frac{2\ell-2i+p+2}{2}\right)} & - \frac{\Gamma(i+\frac{\ell}{2})\Gamma(\ell-1+2+\frac{\ell}{2})}{(i-1)!(\ell-1)\Gamma\left(\frac{21+p}{2}\right)\Gamma\left(\frac{2\ell+2+p-2i}{2}\right)} \\
\frac{\Gamma(\ell-1+1+\frac{\ell}{2})\Gamma(i+1+\frac{\ell}{2})}{(\ell-1)!(i-1)!\Gamma\left(\frac{2\ell-2i+2+p}{2}\right)\Gamma\left(\frac{21+p}{2}\right)} & + \frac{\Gamma(i+1+2+\frac{\ell}{2})\Gamma(1+\frac{\ell}{2})}{(i-1)!(i-1+1)\Gamma\left(\frac{2\ell-2i+2+p}{2}\right)\Gamma\left(\frac{21+p}{2}\right)}
\end{align*}
\]

Writing the above as \( A + B + A' + B' \), we have

\[
A = - \frac{(\ell-1+1)}{1} A', \quad B = \frac{-1}{\ell-1+1} B'
\]

so that the above is non-negative whenever \( i \leq \frac{\ell+1}{2} \), which covers the range of summation. \( \text{Q.E.D.} \)

2.3 **Practicality of the estimators** \( \varphi_{\varepsilon} \).

Because of the evident difficulty involved in applying any of the estimators (2.1.6) to actual data, their use would not be reasonable unless a significant improvement over

\[
(2.3.1) \quad \varphi(X) = (1 - \frac{p-2}{\|x\|^2})x
\]

resulted. In this subsection evidence will be given to support the use of (2.3.1) as opposed to using one of the formal Bayes estimators (2.1.6).

For \( \varepsilon = 0 \) and \( p \) an even number (2.1.6) simplifies to
\[
\varphi_o(x) = \frac{\frac{1}{2^{rac{1}{2}}} \frac{1}{\Gamma\left(\frac{p+2j+2}{2}\right)} \left(\frac{1}{2}\|x\|^2\right)^j}{\Gamma\left(\frac{p+2j}{2}\right)} x
\]

\[
= \frac{\|x\|^2 e^{-\frac{p-4}{2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}}}{\frac{p-4}{2}} - \frac{\frac{p-2}{2}}{\|x\|^2} \left(\frac{\|x\|^2 e^{-\frac{p-4}{2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}}}{\|x\|^2} - \frac{\frac{p-4}{2}}{\|x\|^2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}\right) x, \quad p=6,8,\ldots
\]

and, when \( p=4 \)

\[
\varphi_o(x) = \left(\frac{\|x\|^2 e^{-\frac{2}{2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}}}{\|x\|^2} - \frac{2}{\frac{g}{2}} \left(\frac{\|x\|^2 e^{-\frac{2}{2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}}}{\|x\|^2} - 1\right)\right) \cdot x.
\]

We now wish to show that \( \varphi_o(x) \) is closely related to (2.3.1).

Writing

\[
D_p(\|x\|)^2 = e^{-\frac{2}{2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!}} - \frac{\frac{p-4}{2}}{\|x\|^2} \frac{(\frac{1}{2}\|x\|^2)^i}{i!},
\]

(2.3.2) becomes

\[
D_p(\|x\|^2) + \frac{\frac{1}{2}\|x\|^2}{\|x\|^2} - \frac{\frac{p-2}{2}}{\|x\|^2} D_p(\|x\|^2)
\]

\[
\frac{\|x\|^2}{D_p(\|x\|)^2} \cdot x
\]

\[
= (1 - \frac{\frac{p-2}{2}}{\|x\|^2}) x + \frac{\frac{1}{2}\|x\|^2}{D_p(\|x\|)^2} \frac{\frac{p-4}{2}}{\|x\|^2} \cdot x
\]

14
which, for even moderately large values of $\|x\|^2$ is approximately equal to $(1 - \frac{p-2}{\|x\|^2})x$, e.g., the estimator given by (2.3.1).

From the above computation it is evident that if $\varphi_o(x)$ is to offer any substantial improvement over (2.3.1) it must do so for small values of $\|x\|^2$, i.e., for small values of $\|\theta\|^2$, since $\mathbb{E}\|x\|^2 = \|\theta\|^2 + p$. But it is precisely for small values of $\|\theta\|^2$ that (2.3.1) is most effective, indeed, the risk of (2.3.1) when $\|\theta\|^2 = 0$ is 2 [2]. In section 2.5 another estimator $\varphi^*_o(x)$ will be shown to have everywhere lower risk than (2.3.1) and to be, computationally, as simple as (2.3.1).

2.4 The estimators of James and Stein.

In [2] it was shown that, for $p \geq 3$, each of the estimators

$$\varphi_c(x) = (1 - \frac{c}{\|x\|^2})x, \quad 0 < c < 2(p-2),$$

of the mean of a $p$-variate multivariate normal distribution with known covariance matrix $I_p$, has everywhere lower risk than $\varphi_o(x) = x$.

Here the loss function is the usual sum of mean-square errors

$$L(\varphi(x), \theta) = \sum (\varphi_i(x) - \theta_i)^2,$$

where $\varphi_i(X)$ is the $i$th element of the $p \times 1$ vector $\varphi(X)$. Now $\varphi_c(x)$ offers, a fortiori, an estimate of $\theta_i$, $i=1,2,\ldots,p$; specifically this estimate is $\varphi_{c1}(X)$. Since it is known that $X_1$ is a minimax-admissible estimator of $\theta_1$ for squared error loss, $\varphi_c(x)$ must for some
values of $\theta$ be inferior to $X_1$. Clearly, since $\varphi_c(X)$ is better than $X$ in terms of the loss function (2.4.2), any excessive loss in estimating some particular $\theta_1$ will be more than compensated for in the estimates of some of the other components of $\theta$. The natural questions which arise are what is the risk of $\varphi_c(X)$, and how does this risk vary as a function of $\theta$? The following theorem gives the risk of $\varphi_c(X)$ and also yields the information that the proportion of the total risk $(E \Sigma_{\varphi_c(X)}(X_1-\theta_1)^2)$ attributable to $\varphi_c(X)$ increases as $\frac{\theta_1^2}{\|\theta\|^2}$ increases, holding $\|\theta\|^2$ constant.

**Theorem.** Let $X$ be a $p \times 1$ random vector distributed as a multivariate normal with unknown mean $\theta$ and known covariance matrix $I_p$. Then the estimator (2.4.1) has component-wise risk

$$(2.4.3) \quad E_{\theta}((1- \frac{c}{|X|^2})X_1-\theta_1)^2 =$$

$$1 + \frac{\theta_1^2}{\|\theta\|^2} E \frac{8c^2u}{(2u+p)(2u+p-2)} + E \frac{c^2+4c-2pc-4cu}{(2u+p)(2u+p-2)},$$

where $u$ is a Poisson random variable with parameter $\frac{1}{2} \|\theta\|^2$.

**Proof.** $E((1- \frac{c}{|X|^2})X_1-\theta_1)^2 = E(X_1-\theta_1)^2 - 2c E \frac{X_1(X_1-\theta_1)}{|X|^2} + c^2 E \frac{x_1^2}{|X|^4}$.

Computing this term by term we have

(i) $E(X_1-\theta_1)^2 = 1$

(ii) $E \frac{x_1^2}{|X|^4} = E \left\{ \frac{x_1^2}{(x_1^2 + x_{2L+p-1}^2)} \right\}$
where \( K \) has a Poisson distribution with parameter \( \frac{1}{2} \theta_1^2 \) and \( L \) has a Poisson distribution with parameter \( \frac{1}{2} \sum_{j \neq 1} \theta_j^2 \). But this equals

\[
E \left( \frac{1+2k}{p+2k+2L} \right) \cdot \frac{1}{p+2k+2L}
\]

\[
= E \left\{ E \left[ \frac{1+2k}{(p+2k+2L)(p-2+2k+2L)} \mid K+L \right] \right\}
\]

\[
= E \left( \frac{1}{(p+2u)(p-2+2u)} + \frac{\theta_1^2}{\|\theta\|^2} E \left( \frac{2u}{(p+2u)(p-2+2u)} \right) \right)
\]

where \( U = K + L \), because \( E(K \mid K+L) = \frac{\theta_1^2}{\|\theta\|^2} (K+L) \) and \( K \) and \( L \) are independent. \( U \) has a Poisson distribution with parameter \( \|\theta\|^2 \).

(iii) \( E \frac{\chi_1^2}{\|X\|^2} = E \left\{ E \left( \frac{\chi_1^2}{\chi_1^2 + \chi_2^2 \mid p-2+2L} \right) \mid K, L \right\} \)

\[
= E \frac{1+2k}{p+2k+2L}
\]

\[
= E \left( E \left[ \frac{1+2k}{p+2k+2L} \mid K+L \right] \right)
\]

\[
= E \frac{1}{p+2u} + \frac{\theta_1^2}{\|\theta\|^2} E \frac{2u}{p+2u}.
\]

(iv) The computation of \( E \frac{\theta_1 X_1}{\|X\|^2} \) requires a trick:
\[
\begin{align*}
E\left( \frac{x_i}{\|x\|^2} \right) &= \theta_i^2 \frac{1}{\|x\|^2} + E \left( \frac{x_i - \theta_i}{\|x\|^2} \right) \\
&= \theta_i^2 \frac{1}{p-2+2u} + \theta_i \frac{d}{d\theta_i} \left[ \frac{1}{(2\pi)^{1/2}} \int \cdots \int e^{-\frac{1}{2}x_i^2} \frac{1}{\|x\|^2} dx_1 \cdots dx_p \right] \\
&= \theta_i^2 \frac{1}{p-2+2u} + \theta_i \frac{d}{d\theta_i} \left[ e^{-\frac{1}{2}\|\theta\|^2} \left( \frac{\|\theta\|^2}{2} \right)^s \right] \\
&= \theta_i^2 \frac{1}{p-2+2u} + \theta_i \sum_{s=0}^{\infty} \frac{\left( \frac{1}{2} \|\theta\|^2 \right)^s}{s! (p-2+2s)} \\
&= \theta_i^2 \sum_{s=0}^{\infty} \frac{\left( \frac{1}{2} \|\theta\|^2 \right)^s}{s! (p-2+2s)} \\
&= \theta_i^2 \sum_{s=0}^{\infty} \frac{\left( \frac{1}{2} \|\theta\|^2 \right)^{s-1}}{s! (p-2+2s)} \\
&= \frac{\theta_i^2}{\|\theta\|^2} \sum_{s=0}^{\infty} \frac{\left( \frac{1}{2} \|\theta\|^2 \right)^s}{s! (p-2+2s)} \\
\end{align*}
\]

where \( u \) has a Poisson distribution with parameter \( \frac{1}{2} \|\theta\|^2 \).

Combining terms, equation (2.4.3) follows.
2.5 A simple improvement of some estimates.

An obvious objection to the estimator given in (2.4.1) is that the coefficient of $X$, $(1 - \frac{c}{\|X\|^2})$ can be negative. Thus Stein [2] suggests replacing $(1 - \frac{c}{\|X\|^2})$ by $\max(0, 1 - \frac{c}{\|X\|^2})$. That the new estimator

$$(2.5.1) \quad \varphi^*_c(X) = \max(0, 1 - \frac{c}{\|X\|^2}) X$$

has everywhere lower risk than (2.4.1) follows from the theorem to be proved in this subsection.

Suppose a $p$-dimensional parameter vector

$$(2.5.2) \quad \theta = (\theta_1, \theta_2, \ldots, \theta_p)'$$

is estimated by

$$(2.5.3) \quad \varphi(X) = (h_1(X,S)X_1, h_2(X,S)X_2, \ldots, h_p(X,S)X_p)'$$

where $X$ is $p \times 1$ and $S$ is some other statistic. We next define the estimator

$$(2.5.4) \quad \varphi^*(X) = (h^*_1(X,S)X_1, h^*_2(X,S)X_2, \ldots, h^*_p(X,S)X_p)'$$

in which $h^*_i(X,S) = \max(0, h_i(X,S))$, $i = 1, 2, \ldots, p$.

Theorem. With sum of squares loss $\varphi^*$ has everywhere lower risk than $\varphi$ whenever at least one $h_i$ has non-zero probability of being negative and, for such $h_i$

$$\sgn E[X_1|h_i(X,S) < 0] = \sgn \theta_i.$$
It is understood that \( \operatorname{sgn} A = \) the algebraic sign of \( A \).

Proof of the theorem: Let \( \operatorname{P}[h_k(X,S) < 0] > 0 \) and observe the contribution of \( h_k(X,S)X_k \) to the risk:

\[
\mathbb{E} h_k^2(X,S)X_k^2 - 2\theta_k \mathbb{E} h_k(X,S)X_k + \theta_k^2.
\]

We may ignore \( \theta_k^2 \) in the computation which follows, the remaining terms equalling

\[
\mathbb{E} \mathbb{E}(h_k^2(X,S)X_k^2 - 2\theta_k \mathbb{E} h_k(X,S)X_k | h_k(X,S) < 0)
\]

\[
+ \mathbb{E}(h_k^2(X,S)X_k^2 - 2\theta_k \mathbb{E} h_k(X,S)X_k | h_k(X,S) \geq 0).
\]

Since \( \varphi^* \) involves a change in \( \varphi \) only when \( h_k(X,S) < 0 \) we are concerned with just the first member of this sum.

Clearly

\[
\mathbb{E}(\mathbb{E} h_k^2(X,S)X_k^2 | h(X,S) < 0) = 0
\]

\[
< \mathbb{E}(h_k^2(X,S)X_k^2 | h(X,S) < 0)
\]

since \( h(X,S) < 0 \) with positive probability.

By hypothesis

\[
\mathbb{E}((-2\theta_k)X_k | h_k(X,S) < 0) < 0.
\]

Therefore

\[
\mathbb{E}((1-2\theta_k)h_k(X,S)X_k | h_k(X,S) < 0) > 0.
\]

But

\[
\mathbb{E}((-2\theta_k)h_k^2(X,S)X_k | h_k(X,S) < 0) = \mathbb{E} 0 = 0.
\]

This proves the theorem.
In the estimator of James and Stein [2] the condition of the theorem is that $E[\theta_i x_i \|x\|^2 < c] > 0$, which is trivial. Let $a > 0$, $a^2 < c$, then

$$E[x_i \|x\|^2 < c] = E[E[x_i | x_i^2 = a^2] \sum_{i \neq j} x_j^2 < c - a^2]$$

$$= E[aP[x_i = a | x_i^2 = a^2] - aP[x_i = -a | x_i^2 = a^2] \sum_{i \neq j} x_j^2 < c - a^2]$$

$$= E[f_1(a) \sum_{i \neq j} x_j^2 < c - a^2].$$

But $f_1(a)$ has the same sign as $\theta_i$ since, if $a$ is closer to $\theta_i$ than is $-a$, $f(a)$ is positive, and if $-a$ is closer to $\theta_i$ (which means $\theta_i$ is negative), $f(a)$ is negative.

Bhattacharya [1] uses estimators of the form (2.4.1) to construct an improved estimator of the mean of a multivariate normal when the covariance matrix of the distribution is not equal to the inverse of the matrix appearing in the loss function. It is an immediate consequence of his procedure and the result just shown that his estimator would be improved by replacing estimators of the form (2.4.1) by those of type (2.5.1) in constructing his estimator. This result also follows directly from the theorem.
SECTION 3
MULTIPLE REGRESSION

3.1 The problem

In this subsection we shall state an estimation problem of multiple regression, and shall show how the group structure of the problem both enables us to guess a reasonable solution and to compute the risk of that solution.

Suppose $X_1, X_2, \ldots, X_n$ are independent $(p+1)$-dimensional random vectors, each distributed as a multivariate normal with mean $\Theta$ and covariance matrix $\Sigma$. We will use the following partitions

$$X_i = \begin{pmatrix} Y_i \\ Z_i \end{pmatrix}, \quad i = 1, 2, \ldots, n,$$

$$\Theta_i = \begin{pmatrix} \eta \\ \zeta \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} A & B' \\ B & \Gamma \end{pmatrix},$$

where $Y_i$ and $A$ are $1 \times 1$, $Z_i$ and $B$ are $p \times 1$.

It follows that

$$E(Y_i | Z_i) = \alpha + \beta'Z_i,$$

where

$$\beta = \Gamma^{-1}B,$$

$$\alpha = \eta - \beta'\zeta.$$
For the problems of estimating \((\alpha, \beta)\) and of estimating \(\beta\) alone we use the loss functions proposed in [3] which are, respectively,

\[
L((\theta, \xi), (\hat{\theta}, \hat{\xi})) = \frac{[(\hat{\theta}-\alpha)^+ + (\hat{\xi}-\beta)^+]^2 + (\hat{\beta} - \beta)^\top \Gamma(\hat{\beta} - \beta)}{A-B'R^{-1}B}.
\]

(3.1.7)

\[
L((\beta, \xi), \hat{\beta}) = \frac{(\hat{\beta} - \beta)^\top \Gamma(\hat{\beta} - \beta)}{A-B'R^{-1}B}.
\]

(3.1.8)

The maximum likelihood estimators, which have constant risk with the loss functions shown above, are, respectively

\[
\hat{\alpha}_o = \bar{Y} - \hat{\beta}_o' \bar{Z}, \quad \hat{\beta}_o = V^{-1}U,
\]

(3.1.9)

\[
\hat{\beta}_o = V^{-1}U,
\]

(3.1.10)

where

\[
U = \sum_{i=1}^{n} Z_i Y_i - n \bar{Z} \bar{Y},
\]

(3.1.11)

\[
V = \sum_{i=1}^{n} Z_i Z_i' - n \bar{Z} \bar{Z}'.
\]

(3.1.12)

Stein [3] proved that the estimators (3.1.9) and (3.1.10) each have the minimax property with respect to the loss functions (3.1.7) and (3.1.8). Moreover, he also showed that (3.1.9) is admissible when \(p=1\) and \(n \geq 6\), and that (3.1.10) is admissible for \(p=1, n \geq 4\) and for \(p=2, n \geq 6\). In addition [3] contains the result that (3.1.10) is inadmissible when \(p \geq 3\), but does not exhibit a specific estimator which has everywhere lower risk than (3.1.10).
The following transformations [3] leave the problems of estimating $(\alpha, \beta)$ and, hence, of $\beta$ alone, invariant:

\[(3.1.13) \quad \begin{pmatrix} Y_1 \\ Z_1 \end{pmatrix} \rightarrow \begin{pmatrix} aY_1 + b'Z_1 + d \\ cZ_1 + e \end{pmatrix}, \]

\[(3.1.14) \quad \begin{pmatrix} \eta \\ \zeta \end{pmatrix} \rightarrow \begin{pmatrix} a\eta + b'\zeta + d \\ c\zeta + e \end{pmatrix}, \]

\[(3.1.15) \quad \begin{pmatrix} AB' \\ B' \Gamma \end{pmatrix} \rightarrow \begin{pmatrix} a \ b' \\ b \ c' \end{pmatrix} \begin{pmatrix} a \ b' \\ b \ c' \end{pmatrix} \begin{pmatrix} A' \ B \end{pmatrix} \begin{pmatrix} a \ b' \\ b \ c' \end{pmatrix}, \]

\[= \begin{pmatrix} a^2 A + 2ab'B + b'Bb & (ab' + b'c)c' \\ c - \Gamma b & c \Gamma c' \end{pmatrix}, \]

\[(3.1.16) \quad \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \rightarrow \begin{pmatrix} a \hat{\alpha} - ae'c^{-1}b + d - b'c^{-1}e \\ ac^{-1}b + c^{-1}b \end{pmatrix}, \]

where $a$ and $d$ are $1 \times 1$, $b$ and $e$ are $p \times 1$, $c$ is a nonsingular $p \times p$ matrix, $0$ is a vector of $p$ zeros.

Under the subgroup with $b=0$ the estimators

\[(3.1.17) \quad \hat{\alpha} = \bar{X} - \hat{\beta} \bar{Z}, \quad \hat{\beta} = f(\mathcal{R}^2)\mathcal{V}^{-1}U, \]

are invariant, where $f(\mathcal{R}^2)$ is any function of the multiple correlation coefficient

\[(3.1.18) \quad \mathcal{R}^2 = \frac{U'\mathcal{V}^{-1}U}{T}, \]

where

\[(3.1.19) \quad T = \sum_{i=1}^{n} \gamma_i^2 - n \bar{Y}^2. \]

The risk of an estimator of the form (3.1.17) is relatively simple
to compute since it need be computed only at \( \xi = 0, \Gamma = I, A - B' \Gamma^{-1} B = I \).

Then, applying the transformations

\begin{align*}
(3.1.20) \quad & c'c = \Gamma , \\
(3.1.21) \quad & e = c \xi , \\
(3.1.22) \quad & s = (A - B' \Gamma^{-1} B)^{1/2} \end{align*}

we see that the risk we have computed at \((0, I, 1)\) is equal to the risk at \((\xi, \Gamma, A - B' \Gamma^{-1} B)\).

3.2 An everywhere better estimator of \(\beta\), when \(p \geq 3\).

In the preceding section we noted that the maximum likelihood estimator, \(\hat{\beta}_o = V^{-1} U\), is inadmissible when \(p \geq 3\), but, in proving this result, [3] did not exhibit an estimator better than \(\hat{\beta}_o\). We will now proceed to study a sub class of (3.1.17), specifically estimators of the form

\begin{equation}
(3.2.1) \quad \hat{\beta}_c = (1 - c \frac{1-R^2}{R^2}) V^{-1} U = (1 - c \frac{1-R^2}{R^2}) \hat{\beta}_o ,
\end{equation}

for \(c\) a positive constant. It will be shown that, for \(0 < c < \frac{2(p-2)}{n-p+2}\), \(\hat{\beta}_c\) has everywhere lower risk with respect to the loss function (3.1.8) than has \(\hat{\beta}_o\). Within this class of estimators Stein [3] showed that the one corresponding to \(c = \frac{p-2}{n-p+2}\) was best when \(\beta = 0\).

From section 3.1 it is evident that we need only compute the risk of (3.2.1) at \((0, I, 1)\). Therefore, the loss function (3.1.8) becomes

\begin{equation}
(3.2.2) \quad L((\beta, I), \hat{\beta}) = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \|\hat{\beta} - \beta\|^2 .
\end{equation}
Using (3.1.10) and (3.1.18) we have \( \hat{\beta}_c = (1 - c \frac{T_{-\hat{\beta}_o'V \hat{\beta}_o}}{\hat{\beta}_o'V \hat{\beta}_o}) \hat{\beta}_o \) where \( T_{-\hat{\beta}_o'V \hat{\beta}_o} \) is independent of \( \hat{\beta}_o'V \hat{\beta}_o \) and of \( \hat{\beta}_o \), since we have assumed we are computing the risk at \((0, I, 1)\). Letting \( \hat{\beta}_c = \hat{\beta} \), (3.2.2) becomes

\[
L(\hat{\beta}_c, \beta) = \|\hat{\beta}_c - \beta\|^2 - 2c \frac{T_{-\hat{\beta}_o'V \hat{\beta}_o}}{\hat{\beta}_o'V \hat{\beta}_o} \hat{\beta}_o' \hat{\beta}_o (\hat{\beta}_o - \beta) + c^2 \frac{(T_{-\hat{\beta}_o'V \hat{\beta}_o})^2}{(\hat{\beta}_o'V \hat{\beta}_o)^2} \hat{\beta}_o' \hat{\beta}_o.
\]

Taking expectations we have the risk

\[
\rho(\hat{\beta}_c, \beta) = \mathbb{E}\|\hat{\beta}_c - \beta\|^2 - 2c(n-p) \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o (\hat{\beta}_o - \beta)}{\hat{\beta}_o'V \hat{\beta}_o} + c^2(n-p)(n-p+2) \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o}{(\hat{\beta}_o'V \hat{\beta}_o)^2},
\]

because \( T_{-\hat{\beta}_o'V \hat{\beta}_o} \) is \( \chi^2_{n-p} \).

In order to compute the risk (3.2.4) it will prove useful to make a number of transformations. Recalling that \( V \) has the Wishart distribution \( W(p, n; \Sigma_p) \) and that \( \hat{\beta}_o | V \) is normally distributed with mean \( \beta \) and covariance matrix \( V^{-1} \), we let

\[
V = QDQ'
\]

where \( D \) is a diagonal matrix and \( Q \) is orthogonal. If we let

\[
z = Q'V^{-\frac{1}{2}} \hat{\beta}_o = Q' (D^{-\frac{1}{2}}Q') \hat{\beta}_o = D^{-\frac{1}{2}} Q' \hat{\beta}_o,
\]

then \( z | V \) has a normal distribution with mean \( D^{-\frac{1}{2}}Q\beta \) and covariance matrix \( I_p \). We denote the diagonal elements of \( D \) by \( d_1, d_2, \ldots, d_p \).
Using these transformations we have

\[(3.2.6) \quad \hat{\beta}_o \hat{\beta}_o = (QD^{-1}z)'(QD^{-1}z) = z'D^{-1}z, \]

\[(3.2.7) \quad \hat{\beta}_o'\hat{V}\hat{\beta}_o = z'z. \]

We now proceed to compute (3.2.4). In [3] it was shown that, for \( n \geq p + 2 \),

\[(3.2.8) \quad E\|\hat{\beta}_o - \beta\|^2 = \frac{p}{n-p-1}. \]

In order to compute \( E\frac{\hat{\beta}_o}{\hat{\beta}_o'\hat{V}\hat{\beta}_o} \) it is advisable to compute the conditional expectation \( E\left\{ \frac{\beta'\beta}{\beta'V\beta} \left| V \right. \right\} \). This, in terms of the transformations made above, becomes

\[
E\left\{ \frac{(QD^{-1}z)'\beta}{z'D^{-1}z} \left| Q, D \right. \right\} = E\left\{ \frac{z'D^{-1}Q'\beta}{z'D^{-1}z} \left| Q, D \right. \right\}.
\]

This will equal the sum of expectations

\[(3.2.9) \quad E\left\{ \frac{z_i(D^{-1}Q'\beta)_i}{z'D^{-1}z} \left| Q, D \right. \right\}, \quad i = 1, 2, \ldots, p. \]

But (3.2.9) may be written as

\[
(D^{-1}Q'\beta)_i E\left\{ \frac{z_i}{z'D^{-1}z} \left| Q, D \right. \right\} = (D^{-1}Q'\beta)_i \frac{Ez_i}{\Sigma(Ez_i)^2} E\left\{ \frac{2K}{p-2+2K} \left| Q, D \right. \right\}.
\]

where \( K \) has a Poisson distribution with parameter \( \frac{1}{2} EZ_i^2 \), by section 2.4. Now \( Ez_i = (D^{-1}Q'\beta)_i \), so that (3.2.9) is
\[
\frac{1}{(D^2 q' \beta)^2} \left( D^2 q' \beta \right) \frac{1}{E((D^2 q' \beta)^2)}
\]

\[
= \left( \frac{Q' \beta}{\beta' V \beta} \right)^2 \left( \frac{Q' \beta}{\beta' V \beta} \right)^2
\]

\[
E \left( \frac{2K}{p-2+2K} \mid Q, D \right)
\]

Summing on \( i \), we have

\[
E \left( \frac{\hat{\beta}' \beta}{\hat{\beta}' \hat{V}^2 \beta} \mid V \right) = \frac{\beta' \beta}{\beta' V \beta} E \left( \frac{2K}{p-2+2K} \mid V \right)
\]

where \( K \) is Poisson with parameter \( \frac{1}{2} \Sigma (D^2 q' \beta)_i = \frac{1}{2} \beta' V \beta \). Before evaluating \( E \left( \frac{\hat{\beta}' \beta}{\hat{\beta}' \hat{V}^2 \beta} \mid V \right) \), we make an orthogonal transformation which puts "all \( \beta \)" into \( \beta_1 \). Denoting the upper left-hand corner element of \( V \) be \( V_{11} \), we have

\[
\frac{\beta' \beta}{\beta' V \beta} E \left( \frac{2K}{p-2+2K} \mid V \right) = \frac{1}{V_{11}} E \left( \frac{2K}{p-2+2K} \mid V \right)
\]

where the Poisson parameter may now be written as \( \frac{1}{2} \| \beta \|^2 V_{11} \). Therefore,

\[
E \left( \frac{\hat{\beta}' \beta}{\hat{\beta}' V \beta} \right) = E \left( \frac{1}{V_{11}} \right) \sum_{k=0}^{\infty} \frac{e^{-1/2 \| \beta \|^2 V_{11}}}{k! (p-2+2K)} \left( \frac{1/2 \| \beta \|^2 V_{11}}{2k} \right)^k
\]

\[
= E \left( \frac{1}{2} \| \beta \|^2 V_{11} \right) \sum_{k=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{k! (p-2+2K)} \left( \frac{1/2 \| \beta \|^2 V_{11}}{2k} \right)^k
\]

\[
= \sum_{k=0}^{\infty} \frac{2^{k-1} \Gamma(\frac{n}{2} + k-1)}{k! (1 + \| \beta \|^2)^2 \Gamma(\frac{n}{2}) (p-2+2K)}
\]

because
Thus,

\[
E \frac{\|\beta\|^2_{V_{11}}}{2} V_{11}^j = \frac{2^j \Gamma \left( \frac{n}{2} + j \right)}{(1 + \|\beta\|^2)^j + \frac{n}{2} \Gamma \frac{n}{2}} .
\]

The next part of (3.2.4) to be evaluated is \( E \frac{\beta'_o^o \beta_o}{\beta'_o \hat{V}_o^o} \). Proceeding as before we first evaluate the conditional expectation

\[
E \left( \frac{\beta'_o^o \beta_o}{\beta'_o \hat{V}_o^o} | V \right) = E \left( \frac{\beta'_o^o \beta_o}{\beta'_o \hat{V}_o^o} | D, Q \right) = E \left( \frac{\sum d_i}{(\Sigma x_i^2)^2} | D, Q \right) = E \left( \frac{2K_i + 1}{\Sigma d_i} | D, Q \right)
\]

(\text{where } K_i | D, Q \text{ is Poisson with parameter } \frac{1}{2} d_i (Q' \beta)^2_i)

\[
= E \left( \frac{1}{2} \Sigma K_i | D, Q \right) = E \left( \frac{1}{2} \Sigma \frac{d_i}{d_i} + 2(\Sigma K_i) \frac{\Sigma (Q' \beta)^2_i}{\Sigma d_i (Q' \beta)^2_j} | D, Q \right)
\]

\[
= E \left( \frac{1}{(2K + p)(2K + p - 2)} \left[ \text{trace } V^{-1} + 2K \frac{\|\beta\|^2}{\beta' V \beta} \right] | V \right)
\]
where $K$ has a Poisson distribution with parameter $\frac{1}{2} \beta' V\beta$. Making the same orthogonal transformation as before, we obtain

$$E \left( \frac{\hat{\beta}' \hat{\beta}}{(\hat{\beta}' V \hat{\beta})^2} \right) = E \frac{1}{(2K+p)(2K+2)} \left( \text{trace } V^{-1} + \frac{2}{V_{11}} \right),$$

with $K$ being Poisson with parameter $\frac{1}{2} \|\beta\|^2 V_{11}$. Using the fact that $E(\text{trace } V^{-1}|V_{11}) = \frac{1}{V_{11}} \frac{n-2}{n-p-1} + \frac{p-1}{n-p-1}$, we have

$$E \frac{\hat{\beta}' \hat{\beta}}{(\hat{\beta}' V \hat{\beta})^2} = E \frac{1}{(2K+p)(2K+2)} \left( E(\text{trace } V^{-1}|V_{11}) + \frac{2K}{V_{11}} \right)$$

$$= E \frac{1}{(2K+p)(2K+2)} \left[ \left( \frac{n-2}{n-p-1} + 2K \right) \frac{1}{V_{11}} + \frac{p-1}{n-p-1} \right]$$

$$= E \left( \frac{1}{V_{11}} E \left( \frac{1}{(2K+p)(2K+2)} \left( \frac{n-2}{n-p-1} + 2K \right) \left| V_{11} \right| \right) \right) + \frac{p-1}{n-p-1} E \frac{1}{(2K+p)(2K+2)}$$

$$= E \frac{1}{V_{11}} e^{-\frac{1}{2} \|\beta\|^2 V_{11}} \sum_{k=0}^{\infty} \frac{\left( \frac{1}{2} \|\beta\|^2 V_{11} \right)^k}{k!(2k+p)(2k+2)} \left( \frac{n-2}{n-p-1} + 2K \right)$$

$$+ E e^{-\frac{1}{2} \|\beta\|^2 V_{11}} \sum_{k=0}^{\infty} \frac{\left( \frac{1}{2} \|\beta\|^2 V_{11} \right)^k}{k!(2k+p)(2k+2)} \frac{p-1}{n-p-1}$$

$$= \sum_{k=0}^{\infty} \frac{2^{k-1} r^{\frac{n}{2} + k-1} \left( \frac{1}{2} \|\beta\|^2 \right)^k}{(1 + \|\beta\|^2)^{k-1} \frac{n}{2} r^{\frac{n}{2}} k!(2k+p)(2k+2)} \left( \frac{n-2}{n-p-1} + 2K \right)$$

$$+ \sum_{k=0}^{\infty} \frac{2^{k} r^{\frac{n}{2} + k} \left( \frac{1}{2} \|\beta\|^2 \right)^{k+1}}{2(1 + \|\beta\|^2)^{k-1} \frac{n}{2} r^{\frac{n}{2}} k!(2k+p)(2k+2)} \left( \frac{p-1}{n-p-1} \right).$$
So that we have,

\[
(3.2.11) \quad \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o}{(\hat{\beta}_o' \hat{\beta}_o)^2} = \frac{1}{2(1 + \|\beta\|^2)^2 - \frac{n}{\Gamma(n/2)}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{n}{2} + k - 1\right)}{k!} \left(\frac{\|\beta\|^2}{1 + \|\beta\|^2}\right)^k \frac{1}{(2k+p)(2k+p-2)}. 
\]

\[
\left[\frac{n-2}{n-p-1} - 2k + \frac{(n+2k-2)(p-1)}{(1 + \|\beta\|^2)(n-p-1)}\right].
\]

All that is left of (3.2.4) to compute is \( \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o}{\hat{\beta}_o' \hat{\beta}_o} \). But a computation exactly analogous to the evaluation of \( \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o}{(\hat{\beta}_o' \hat{\beta}_o)^2} \) leads us to

\[
(3.2.12) \quad \mathbb{E} \frac{\hat{\beta}_o' \hat{\beta}_o}{\hat{\beta}_o' \hat{\beta}_o} = \frac{1}{2(1 + \|\beta\|^2)^2 - \frac{n}{\Gamma(n/2)}} \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{n}{2} + k - 1\right)}{k!} \left(\frac{\|\beta\|^2}{1 + \|\beta\|^2}\right)^k \frac{1}{2k+p}. 
\]

\[
\left[\frac{n-2}{n-p-1} + 2k + \frac{(n+2k-2)(p-1)}{(1 + \|\beta\|^2)(n-p-1)}\right].
\]

To show that \( \hat{\beta}_c \) is, for some values of \( c \), better than \( \hat{\beta}_o \) it is necessary to show that \( \rho(\hat{\beta}_c, \beta) \leq \rho(\hat{\beta}_o, \beta) \) for all \( \beta \), with strict inequality for some value of \( \beta \). Using the formula (3.2.4) for \( \rho(\hat{\beta}_c, \beta) \), simple algebra leads one to conclude that \( \hat{\beta}_c \) will be better than \( \hat{\beta}_o \) for all values of...
\[(3.2.13) \quad c \leq \frac{2(n-p)}{(n-p)(n-p-2)} E \frac{\hat{\beta}'(\beta - \beta)}{\hat{\beta}' \beta} / E \frac{\hat{\beta}' \beta}{(\hat{\beta}' \beta)^2} . \]

It will now be shown that \( c = \frac{2(p-2)}{n-p+2} \) (and hence all positive values of \( c \) less than this) satisfies (3.2.13). To prove this we must show

\[(3.2.14) \quad (p-2) E \frac{\hat{\beta}' \beta}{(\hat{\beta}' \beta)^2} - E \frac{\beta'(\beta')}{(\hat{\beta}' \beta)^2} \leq 0. \]

Setting \( \lambda = \frac{\|\beta\|^2}{1 + \|\beta\|^2} \) and

\[C_k = \frac{\Gamma\left(\frac{n}{2} + k - 1\right) \lambda^k}{k!(p+2k)(p+2k-2)(1 + \|\beta\|^2)(n-p-1)} \]

(3.2.14) becomes

\[2(p-1) \sum_{k=0}^{\infty} C_k (n-p-2)(1 + \|\beta\|^2) - (n+2k-2) \]

which will be negative whenever

\[(3.2.15) \quad (n-p-2)(1 + \|\beta\|^2) \sum_{k=0}^{\infty} \frac{\Gamma\left(n+2k-2\right) \lambda^k}{k!(2k+2)(2k+p-2)} - \sum_{k=0}^{\infty} \frac{\Gamma\left(n+2k-2\right) \lambda^k}{k!(2k+p)(2k+p-2)} \]

is negative. Letting \( j = k-1 \), proving (3.2.15) is negative reduces to proving the negativity of

\[(3.2.16) \quad \frac{1}{1-\lambda} (n-p-2) \sum_{j=0}^{\infty} \frac{\Gamma\left(n+2j\right) \lambda^j}{j!(p+2j)(p+2j+2)} - 2 \sum_{j=0}^{\infty} \frac{\Gamma\left(n+2j+2\right) \lambda^j}{j!(p+2j)(p+2j+2)} ,\]

32
where we recall that \( \frac{1}{1-\lambda} = 1 + \|\beta\|^2 \).

Multiplying (3.2.16) through by \( 1-\lambda \) and collecting coefficients
of \( \lambda^j \) we have

\[
\sum_{j=0}^{\infty} \frac{\lambda^j \Gamma\left(\frac{n+2j}{2}\right)}{j!(2j+p)(2j+p+2)} \left[ \frac{2^{p+2j+2}}{p+2j-2} j + n - p - 2 - n - 2j \right]
\]

\[= \sum_{j=0}^{\infty} \frac{\lambda^j \Gamma\left(\frac{n+2j}{2}\right)}{j!(2j+p)(2j+p+2)} \left[ \frac{2^{p+2j+2}}{p+2j-2} \left( \frac{2^{j}}{p+2j-2} - 1 \right) \right]
\]

which is negative since each term in the series is negative whenever
\( p \geq 3 \).

For a given value of \( \|\beta\| \) the value of \( c \) which minimizes the risk
\( \rho(\hat{\beta}_c, \beta) \) is given by taking one half the maximum value of \( c \) satisfying
the inequality (3.2.13). Calling this \( c = c(\|\beta\|) \), the computation just
completed demonstrates that \( c(\|\beta\|) \geq \frac{p-2}{n-p+2} \) for all \( \beta \). Unfortunately
this is not an equality, Stein having shown in [3] that \( c(0) = \frac{p-2}{n-p+2} \)
but \( \lim_{\|\beta\| \to \infty} c(\|\beta\|) = \frac{p-2}{n-p+2} \).

3.3 An everywhere better estimator of \((\alpha, \beta)\) when \( p \geq 3 \).

In order to prove the inadmissibility of the maximum likelihood
estimator \((\hat{\alpha}_o, \hat{\beta}_o)\) (3.1.9) of \((\alpha, \beta)\) with respect to the loss function
(3.1.7) it will be shown that a class of estimators, each of the form
(3.1.17), has everywhere lower risk than the risk of \((\hat{\alpha}_o, \hat{\beta}_o)\). This
class of estimators is

\[
(3.3.1) \quad \begin{cases}
\alpha_c = \bar{Y} - \hat{\beta}_c \bar{Z}, \\
\hat{\beta}_c = (1 - c \frac{1-R^2}{R^2}) V^{-1}U \quad \text{for} \quad p \geq 3,
\end{cases}
\]
where

\begin{equation}
0 < c < \frac{2(p-2)}{n-p+2}.
\end{equation}

As demonstrated in section 3.1, we may evaluate the risk of each of these estimators while fixing $\xi=0, \Gamma=I$, and $A-B^{-1}B = 1$. The loss function is thus simplified to

\begin{equation}
L((\alpha, \beta), (\hat{\alpha}_c, \hat{\beta}_c)) = (\hat{\alpha}_c - \alpha)^2 + \|\hat{\beta}_c - \beta\|^2.
\end{equation}

Setting $c \frac{1-R^2}{R^2} = f(R^2)$ and $g(R^2) = 1-f(R^2)$, the risk becomes

\begin{align*}
E[(\bar{Y} - f(R^2)U'V^{-1}\bar{Z}-\alpha)^2 + \|f(R^2)V^{-1}U-\beta\|^2] \\
= E(\bar{Y}-U'V^{-1}\bar{Z}-\alpha)^2 + E\|V^{-1}U-\beta\|^2 \\
+ E(g(R^2)U'V^{-1}\bar{Z})^2 + E\|g(R^2)V^{-1}U\|^2 \\
+ E[g(R^2)U'V^{-1}\bar{Z}(\bar{Y}-U'V^{-1}\bar{Z}-\alpha) + g(R^2)U'V^{-1}(\beta-V^{-1}U)] \\
= \text{Risk}(\hat{\alpha}_c, \hat{\beta}_c) + E g(R^2)U'V^{-1}(I+\bar{Z}\bar{Z}')V^{-1}U \\
+ 2E[E g(R^2)U'V^{-1}[\beta-V^{-1}U+\bar{Z}(\bar{Y}-U'V^{-1}\bar{Z}-\alpha)]|\bar{Z}].
\end{align*}

But the last term is equal to

\begin{align*}
2E g(R^2)U'V^{-1}[\beta-V^{-1}U+\bar{Z}\bar{Z}'(\beta-V^{-1}U)] \\
= 2E g(R^2)U'V^{-1}(I+\bar{Z}\bar{Z}')(\beta-V^{-1}U).
\end{align*}

Therefore,
\[ \rho((\hat{\alpha}_o, \hat{\beta}_o), (\alpha, \beta)) = \rho((\hat{\alpha}_c, \hat{\beta}_c), (\alpha, \beta)) \]
\[ + E g(R^2)u^1 v^{n-1} \sum (2\beta - (2-g(R^2))v^{-1}u) \]
\[ + E g(R^2)u^1 v^{n-1} \sum z' (2\beta - (2-g(R^2))v^{-1}u) \]
\[ = \rho((\hat{\alpha}_o, \hat{\beta}_o), (\alpha, \beta)) \]
\[ + \frac{n+1}{n} E g(R^2)u^1 v^{n-1} (2\beta - (2-g(R^2))v^{-1}u) \]
\[ = \rho((\hat{\alpha}_o, \hat{\beta}_o), (\alpha, \beta)) \]
\[ + \frac{n+1}{n} E g(R^2)u^1 v^{n-1} (\beta - v^{-1}u) + \|g(R^2)v^{-1}u\|^2 \]
\[ = \rho((\hat{\alpha}_o, \hat{\beta}_o), (\alpha, \beta)) + \frac{n+1}{n} \left[ E \|\hat{\beta}_c - \beta\|^2 - E \|\hat{\beta}_o - \beta\|^2 \right] \]
\[ < \rho((\hat{\alpha}_o, \hat{\beta}_o), (\alpha, \beta)) \]

whenever \( c < c < \frac{2(p-2)}{n-p+2} \), by the result of section 3.2. Thus each of the estimators given by (3.3.1) and (3.3.2) has everywhere lower risk than the maximum likelihood estimator. Since it was shown in [3] that \((\hat{\alpha}_o, \hat{\beta}_o)\) has the minimax property, this property also holds for the new estimators.

3.4 Inadmissibility of the usual estimator of \( \alpha \) when \( p \geq 3 \).

A special problem of interest is that of estimating \( \alpha \) when the mean of the predictor variables is known (without loss of generality we shall take \( \xi = 0 \)). In this situation \( \alpha \) is reduced to \( \eta \), the mean of \( Y \).
For a loss function we take

\[(3.4.1) \quad L(\hat{\alpha}, \alpha) = (\hat{\alpha} - \alpha)^2 / A - B' \Gamma^{-1} B .\]

The maximum likelihood estimator is

\[(3.4.2) \quad \hat{\alpha}_o = \frac{\bar{Y}}{U' V^{-1} Z} , \]

which, for the loss function (3.4.2) has the minimax property. This follows from \(\hat{\beta}_o\) and \((\hat{\alpha}_o, \hat{\beta}_o)\) each having the minimax property and also being constant risk estimators.

A proof identical to the proof given in section 3.3 leads to the result that each estimate of the form

\[(3.4.3) \quad \hat{\alpha}_c = \bar{Y} - (1 - c \frac{1 - R^2}{R^2}) U' V^{-1} Z , \quad 0 < c < \frac{2(p-2)}{n+p+2} , \]

has everywhere lower risk than \(\hat{\alpha}_o\), establishing the inadmissibility of \(\hat{\alpha}_o\).

3.5 Recommended estimators for \(\alpha, \beta, (\alpha, \beta)\) when \(p \geq 3\).

A reasonable objection to the estimators computed in the preceding subsections is that they all involve multiplying \(V^{-1} U\) by \((1 - c \frac{1 - R^2}{R^2})\), a factor which may be negative. Following the result obtained in section 2.5 it seems plausible to recommend the following estimators

\[(3.5.1) \quad \hat{\beta} = \max(0, 1 - c \frac{1 - R^2}{R^2}) V^{-1} U , \]

\[(3.5.2) \quad \hat{\alpha} = \bar{Y} - \hat{\beta} V Z , \quad \hat{\beta} = \max(0, 1 - c \frac{1 - R^2}{R^2}) V^{-1} U , \]

36
(3.5.3) \[ \hat{\alpha} = \bar{y} - \max(0, 1 - c \frac{1-R^2}{R^2}) U'^{-1} Z \]

As regards the choice of \( c \), any value of \( c \) between 0 and \( \frac{2(p-2)}{n-p+2} \) would give an improvement over the respective maximum likelihood estimators. However, inasmuch as the optimum value of \( c \) is at least \( \frac{p-2}{n-p+2} \), it would be more reasonable to use a value of \( c \) between \( \frac{p-2}{n-p+2} \) and \( \frac{2(p-2)}{n-p+2} \), according to one's judgment, made before seeing the data, of the approximate value of \( \| \beta \|^2 \).

If, however, one has some prior idea regarding \( \beta \) which can be expressed as "\( \beta_1, \beta_2, \ldots, \beta_{\ell} \) are likely to be substantially larger than \( \beta_{\ell+1}, \ldots, \beta_p \)" then it may be advisable to use a procedure which estimates \( \beta_1, \ldots, \beta_{\ell} \) in the usual way, while estimating \( \beta_{\ell+1}, \ldots, \beta_p \) in the manner given in formula (3.5.1). For example, suppose one of the \( \beta_i \), say \( \beta^* \) is quite large compared to the other \( \beta_1 \). Then the result proved in section 2.4 suggests that if we estimated \( \beta \) by formula (3.5.1) the contribution to the error made by \( \hat{\beta}^* \) (the component of \( \hat{\beta} \) which estimates \( \beta^* \)) might be substantially greater than that made by \( \hat{\beta}^o \) (the component of \( \hat{\beta} \) which estimates \( \beta^* \)). To show that such a "mixing" of estimators would indeed yield a better one would require the computation of the risk of such an estimator. Unfortunately such estimators do not possess the invariance property (section 3.1) which enabled us to compute risks in a relatively simple manner.

From the summary of previous results given in section 3.1 it is evident that the question of admissibility of \( (\hat{\alpha}^o, \hat{\beta}^o) \) when the number of predictors, \( p=2 \) is still open. It is clear that the procedures used in section 3 will not be of use in this case, since, in general,
the optimum coefficient of \( \frac{1-R^2}{R^2} \) is a multiple of \( p-2 \). Moreover, it appears that one must consider estimators of a more complicated form than \( f(R^2) \hat{\beta}_0 \) if there is a counterexample to be found.


<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>City, State</th>
<th>Phone</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Stephen G. Allen</td>
<td>Stanford Research Institute 333 Ravenswood Avenue Menlo Park, California 10</td>
<td>Stanford, California</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Fred C. Andrews</td>
<td>Department of Mathematics</td>
<td>University of Oregon Eugene, Oregon 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr. J. Berkowitz</td>
<td>Quality Control Supervisor</td>
<td>RXDEN Division of United Aircraft Corp. Ketye Department – Helen Street Norwalk, Connecticut 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Julian H. Bigelow</td>
<td>School of Mathematics</td>
<td>Institute for Advanced Study Princeton, New Jersey 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Z. W. Birmbaum</td>
<td>Laboratory of Statistical Research</td>
<td>Department of Mathematics University of Washington Seattle, Washington 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Julius R. Blum</td>
<td>Sandra Corporation</td>
<td>Albuquerque, New Mexico 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Paul Blank</td>
<td>Box 532</td>
<td>Fair Oaks, California 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor C. E. P. Box</td>
<td>Department of Statistics</td>
<td>University of Wisconsin Madison, Wisconsin 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Ralph A. Bradley</td>
<td>Department of Statistics</td>
<td>Florida State University Tallahassee, Florida 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor W. C. Cochran</td>
<td>Department of Statistics</td>
<td>Harvard University 2 Divinity Avenue, Room 311 Cambridge, Massachusetts 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Joseph Daly</td>
<td>Bureau of the Census</td>
<td>Washington 25, D. C. 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Lee Cronbach</td>
<td>Bureau of Educational Research</td>
<td>1067 S. Wright Champaign, Illinois 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Don Davis</td>
<td>Dept. of Industrial Engineering</td>
<td>Columbia University New York, New York 27, New York 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Francis Dreisch</td>
<td>Stanford Research Institute</td>
<td>333 Ravenswood Avenue Menlo Park, California 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr. John A. Dutton</td>
<td>Department of Meteorology</td>
<td>University of Wisconsin Madison, Wisconsin 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Meyerswiss</td>
<td>Department of Mathematics</td>
<td>Northwestern University Evanston, Illinois 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor W. T. Federer</td>
<td>Cornell University</td>
<td>New York, New York 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>George Fishman</td>
<td>Logistics Department</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rand Corporation</td>
<td>1700 Main Street</td>
<td>Santa Monica, California 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Merrill M. Flood</td>
<td>Montclair Health Research Institute</td>
<td>205 North Forest Avenue Ann Arbor, Michigan 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Donald P. Gaver</td>
<td>Westmoreland Research Labs.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Murray A. Gelsler</td>
<td>Logistics Section</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The RAM Corporation</td>
<td>1700 Main Street</td>
<td>Santa Monica, California 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor M. P. Goode</td>
<td>Dept. of Industrial Engineering</td>
<td>Administration Cornell University Ithaca, New York 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Theodore E. Harris</td>
<td>The RAM Corporation</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor W. Hiensch</td>
<td>Institute of Mathematical Sciences</td>
<td>New York University New York, New York 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Harold Hotelling</td>
<td>Associate Director</td>
<td>Institute of Statistics 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of North Carolina</td>
<td>Chapel Hill, North Carolina</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Stanley Isaacson</td>
<td>4706 Lakeview Drive</td>
<td>Des Moines 12, Iowa 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Oscar Kemphorne</td>
<td>Statistics Laboratory</td>
<td>Iowa State College Ames, Iowa 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor Solomon Kullback</td>
<td>Department of Statistics</td>
<td>George Washington University Washington 7, D. C. 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Eugene Lukacs</td>
<td>Department of Mathematics</td>
<td>Catholic University Washington 17, D. C. 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor G. W. McElrath</td>
<td>Department of Mechanical Engineering</td>
<td>University of Minnesota Minneapolis 14, Minnesota 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Professor Paul Meier  
Department of Statistics  
University of Chicago  
Chicago, Illinois

Professor James M. Moore  
Dept. of Industrial Engineering  
Northeastern University  
Boston 12, Massachusetts

D. E. Newhall  
Chief, Industrial Eng., Division  
Computrol  
Hdqs., San Bernardino Air  
Material Area  
Norton Air Force Base, California

Professor J. Neyman  
Department of Statistics  
University of California  
Berkeley 4, California

Mr. Edward Paulson  
72-10 43rd Avenue  
Woodside 77  
New York, New York

Mr. W. J. Peterson  
Dept. 65-24  
Lockheed Missile Systems Division  
Sunnyvale, California

Dr. Richard Post  
Department of Mathematics  
San Jose State College  
San Jose, California

Professor Ronald Pyke  
Department of Mathematics  
University of Washington  
Seattle 5, Washington

Professor George J. Resnikoff  
Department of Industrial Engineering  
Illinois Institute of Technology  
Chicago 16, Illinois

Professor Herbert Robbins  
Mathematical Statistics Dept.,  
Columbia University  
New York 27, New York

Professor Murray Rosenblatt  
Department of Mathematics  
Brown University  
Providence 32, Rhode Island

Professor Herman Rubin  
Department of Statistics  
Michigan State University  
East Lansing, Michigan

Professor I. R. Savage  
Department of Statistics  
University of Minnesota  
Minneapolis, Minnesota

Professor George J. Schick  
Department of Mathematics & Statistics  
Sacramento State College  
Sacramento, California

Professor Seymour Sherman  
Department of Mathematics  
Wayne State University  
Detroit 2, Michigan

Professor W. L. Smith  
Department of Statistics  
University of North Carolina  
Chapel Hill, North Carolina

Dr. Milton Sobel  
Department of Statistics  
University of Minnesota  
Minneapolis, Minnesota

Professor Frank Spitzer  
Department of Mathematics  
Cornell University  
Ithaca, New York

Professor Gerald L. Thompson  
Carnegie Institute of Technology  
Schenley Park  
Pittsburgh 33, Pennsylvania

Professor Donald Truax  
Department of Mathematics  
University of Oregon  
Eugene, Oregon

Mr. Harry Weingarten  
U.S. Arms Control and  
Disarmament Agency  
State Dept. Bldg.,  
21st St. and Virginia Ave., N.W.,  
Washington 25, D.C.

Capt. Burton L. Weller  
AFPR Office  
Martin Aircraft Corporation  
Denver, Colorado

Professor M. B. Wilk  
Statistics Center  
Rutgers-The State University  
New Brunswick, New Jersey

Dr. John D. Wilkes  
Office of Naval Research  
Code 200  
Washington 25, D.C.

Professor S. S. Wilks  
Department of Mathematics  
Princeton University  
Princeton, New Jersey

Professor J. Woldoffitz  
Department of Mathematics  
Cornell University  
Ithaca, New York

Professor M. A. Woodbury  
Department of Mathematics  
New York University  
New York 53, New York

Distribution via ONR London

Commodating Officer  
Branch Office  
Navy No. 100  
Fleet Post Office  
New York, New York

Dr. Dennis V. Lindley  
Statistical Laboratory  
University of Cambridge  
Cambridge, ENGLAND

Other Foreign Addresses

Professor J. E. Moyal  
Institute of Advanced Studies  
Australian National University  
Canberra A.C.T.,  
AUSTRALIA

Dr. A. R. Roy  
Statistical Wing  
Indian Council of Agricultural Research  
Llinlithgow Avenue  
New Delhi, INDIA

Professor Tozio Kitagawa  
Mathematical Institute  
Faculty of Science  
Kyoto University  
Fukuoka, JAPAN

Mr. Cesarea Villegas  
Instituto de Matematica y  
Estadistica  
Av. J. Herrera y Resig 565  
Montevideo, URUGUAY

Professor G. S. Watson  
Department of Mathematics  
University of Toronto  
Toronto 5, Ontario,  
CANADA

Professor D. A. S. Fraser  
Department of Mathematics  
University of Toronto  
Toronto 5, Ontario,  
CANADA

Professor H. Witting  
Institut fur Mathematische Statistik  
University of Munster  
Schlossplatz 2  
44 Munster, Westf.,  
GERMANY

Mr. Geoffrey Gregory  
10 Morris Avenue  
Box Hill North  
E.12 Victoria  
AUSTRALIA

Additional Coeoes for Project Leader  
and Assistants, and Reserve for  
Future Requirements  

Contract No. 225(21)  
April 1963