HOTELLING'S $t^2$ TEST UNDER NON-NORMAL CONDITIONS

BY

M. L. EATON and B. EFRON

TECHNICAL REPORT NO. 11
AUGUST 22, 1968

SUPPORTED BY PUBLIC HEALTH SERVICE
GRANT USPHS-GM-14554-02

and

NATIONAL SCIENCE FOUNDATION
GRANT NO. GP-6681

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
HOTELLING'S $T^2$ TEST UNDER NON-NORMAL CONDITIONS

BY

M.L. Eaton and B. Efron

TECHNICAL REPORT NO. 11
AUGUST 22, 1968

SUPPORTED BY PUBLIC HEALTH SERVICE

GRANT USPHS-GM-14554-02

and

NSF GP-6681

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
HOTELLING'S $T^2$ TEST UNDER NON-NORMAL CONDITIONS

By

M.L. Eaton
University of Chicago

B. Efron
Stanford University

1. Hotelling's One-Sample $T^2$ Test.

Multivariate data is often represented by an $n \times k$ data matrix

$$X = \begin{pmatrix}
  x_{11} & x_{12} & \cdots & x_{1k} \\
  \vdots & \vdots & & \vdots \\
  x_{i1} & \cdots & \cdots & x_{ij} \\
  \vdots & \vdots & & \vdots \\
  x_{n1} & \cdots & \cdots & x_{nk}
\end{pmatrix}$$

where each row records $k$ separate measurements made on one of a sample of $n$ individuals. Under the familiar assumption that the rows are independent observations from the same multivariate normal distribution,

$$(x_{i1}, x_{i2}, \ldots, x_{ik}) \overset{iid}{\sim} N(\mu, \Sigma), \quad i=1,2,\ldots,n,$$

Hotelling's one-sample $T^2$ statistic

$$T^2 = \frac{\bar{x}'(S_n)^{-1} \bar{x}}{k}$$

is commonly used to test the null hypothesis

$$H_0: \mu = 0.$$ 

Here $\bar{x}$ and $S_n$ are the sample mean vector and covariance matrix,

$$\bar{x} = \frac{1}{n}X, \quad S = X'X - n\bar{x}'\bar{x}.$$
(where we have defined
\[
\mathbf{e} = \left( \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}} \right),
\]
the unit main diagonal in Euclidean n-space $\mathbb{E}^n$.)

Under the null hypothesis $\mu = 0$ it is easy to show [1, p. 106] that $\frac{n-k}{k} T^2$ has a standard $F_{k,n-k}$ distribution. A quick geometric proof of this fact can be sketched out as follows: standard algebraic manipulations show that

\[
T^2 = \cot^2 \Theta,
\]

where $\Theta$ is the angle between $\mathbf{e}$ and $\mathbf{f}(X)$, the $k$-dimensional linear subspace of $\mathbb{E}^n$ generated by the columns of $X$. Under $H_0$, $\mathbf{f}(X)$ has a uniform random orientation in $\mathbb{E}^n$, which can be seen from the fact that $\mathbf{f}X$ has the same distribution as $X$ for any $n \times n$ orthogonal matrix $\mathbf{F}$.

To compute the distribution of $\cot^2 \Theta$, it is equivalent to think of $\mathbf{f}(X)$ as fixed, say as the subspace generated by the first $k$ coordinate axes, and $\mathbf{e}$ as being any random vector whose direction is distributed uniformly on the unit sphere in $\mathbb{E}^n$. In particular, we can take the components of $\mathbf{e}$ to be independent $N(0,1)$ variates, $e_i \overset{\text{iid}}{\sim} N(0,1)$ for $i=1,2,\ldots,n$, and then $\cot^2 \Theta = (e_1^2 + \cdots + e_k^2)/(e_{k+1}^2 + \cdots + e_n^2)$ is by definition a $\frac{k}{n-k} F_{k,n-k}$ random variable. (This is essentially Hotelling's original proof [5].)
2. Purpose of This Paper.

The assumption of independent \( N(0, \mathbf{I}) \) rows for \( X \) is essential to the above derivation, since it guarantees that \( \mathbf{I}(X) \) will have a uniform random orientation in \( \mathbb{E}^n \). Unfortunately this pleasant symmetry property almost characterizes the \( N(0, \mathbf{I}) \) distribution. Under most other sampling situations where we like to accept the null hypothesis \( H_0 \), the \( T^2 \) statistic will not have an \( F \) distribution.

The purpose of this paper is to discuss the \( T^2 \) statistic under a much weaker symmetry condition, which is satisfied under the null hypotheses of many standard sampling situations.

**Definition:** The random matrix \( X \) is said to have **ORTHANT SYMMETRY** if its distribution is invariant under all sign changes of the rows (i.e., if \( X \) has the same distribution as \( DX \) for any choice of the diagonal matrix \( D \) with diagonal elements \( \pm 1 \)).

In particular, orthant symmetry holds whenever the rows of \( X \) are independent and symmetrically distributed about the origin. It is not necessary that the rows be identically distributed. Thus we could have independent \( N(0, \mathbf{I}_i) \) rows, where the \( \mathbf{I}_i \) are all different. Cases of orthant symmetry with dependent rows arise naturally in considering "nonparametric" versions of the \( T^2 \) statistic (see section 5.)

3. The \( T^2 \) Statistic Under Orthant Symmetry.

We prefer to work with \( T^2 \) in the equivalent form

\[
R^2 = \frac{T^2}{1 + T^2}.
\]
By elementary trigonometry,

\[ R^2 = \cos^2 \Theta, \]

where, as before, \( \Theta \) is the angle between \( e \) and \( \mathcal{L}(X) \). To avoid

trivialities, we assume in what follows that the probability of a zero

row is zero. We will show that under the null hypothesis of ortant

symmetry the distribution of \( nR^2 \) can be approximated by a \( X_k^2 \) distri-

bution, and that this approximation is always conservative in a sense

that will be made explicit below.

For any given matrix \( X \), let us define

\[
\mathcal{D}(X) = \left\{ DX|D = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_n \end{pmatrix}, \delta_i = \pm 1, i=1,2,\ldots,n \right\},
\]

so that \( \mathcal{D}(X) \) is the set of all \( 2^n \) matrices obtainable from \( X \) via

sign changes of the rows. By orthant symmetry we have

\[
P[\bar{x}|\mathcal{D}(X)] = \frac{1}{2^n} \text{ for every } \bar{x} \in \mathcal{D}(X).
\]

Suppose that \( C \) is a basis for \( \mathcal{L}(X) \); that is, \( C \) is an \( n \times k \)

orthonormal matrix whose columns span \( \mathcal{L}(X) \):

\[
\mathcal{L}(C) = \mathcal{L}(X), \ C'C = I.
\]

(We are assuming full rank for \( X \). If the columns are linearly depend-

dent the following arguments hold with \( k \) reduced to the dimension

of \( \mathcal{L}(X) \).) We see that
\[ nR^2 = n \cos^2 \Theta = \sqrt{n} \epsilon C'e' \sqrt{n} . \]

If \( \tilde{X} = DX \) is any other member of \( \mathcal{B}(X) \), then \( \tilde{C} = DC \) is a basis for \( I(\tilde{X}) \) and the new value of our statistic is

\[ nR^2 = \sqrt{n} eDC C'De'\sqrt{n} . \]

Conditional on our observed matrix belonging to the set \( \mathcal{B}(X) \), all \( 2^n \) possible choices of \( D \) are equally likely, yielding \( 2^n \) equally likely realizations of \( n \cos^2 \Theta \). We express this symmetry in a simple but useful lemma:

**Lemma.** Given \( \mathcal{B}(X) \), \( nR^2 \) has conditional distribution \( \Delta CC'\Delta' \), where \( \Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n) \) is a random vector whose components are independent and take values \(+1\) or \(-1\) with probabilities \( \frac{1}{2} \), and \( C \) is a basis for any member of \( \mathcal{B}(X) \).

The k-component random vector

\[ V = \Delta C \]

has, conditional on \( \mathcal{B}(X) \), expectation 0 and covariance matrix I:

\[ \mathbb{E}[V|\mathcal{B}(X)] = (\Delta C) C = 0, \]
\[ \mathbb{E}[V'V|\mathcal{B}(X)] = C'(\Delta \Delta') C \]
\[ = I . \]

It is the sum of \( n \) independent though not identically distributed random vectors, \( V = \sum_{i=1}^{n} \Delta_i (c_{i1}, c_{i2}, \ldots, c_{ik}) \), and the central limit theorem suggests an approximately normal distribution for \( V \),
\[ V \sim N_k(0, I) . \]

The validity of this approximation in the univariate case \( k=1 \) has been examined in a previous paper by one of the authors [4]. The main theorem of that paper can be applied here to show that for any \( k \times 1 \) column vector \( a \), and for any even power \( v \), orthant symmetry implies that

\[ \mathcal{E}(Va)^v \leq \|a\|^v \mathcal{E}_N^v(0,1) , \]

with equality for \( v = 2 \) and strict inequality for \( v = 4, 6, 8, \ldots \).

In this sense the normal approximation to the distribution of \( V \) is conservative in that it overestimates the higher even moments of any linear combination of \( V \)'s components.

Here we are interested in a quadratic form of \( V \),

\[ VV' = nR^2 . \]

The normal approximation suggests a \( \chi_k^2 \) distribution for this quantity. Our main result shows that this approximation is also conservative:

**Theorem:** Under orthant symmetry

\[ \mathcal{E}(nR^2)^p \leq \mathcal{E}(\chi_k^2)^p \]

for any positive integer power \( p \), with equality for \( p = 1 \) and strict inequality for \( p = 2, 3, 4, \ldots \).

By the lemma, it is sufficient to verify \( E[(VV')^p|x(X)] \leq E(\chi_k^2)^p \) for every possible realization of \( P(X) \). This is done in section 6, and some indication of the magnitude of \( \mathcal{E}(\chi_k^2)^p - \mathcal{E}(nR^2)^p \) is given for the cases \( p = 2 \) and \( p = 3 \).
4. **Using the $T^2$ Test.**

To use Hotelling's $T^2$ to test the null hypothesis of orthant symmetry we suggest computing $nR^2$ either from $T^2$ or directly from

$$nR^2 = n\overline{x^2} - \frac{1}{\overline{x}}$$

$$= n\overline{x^2} + \frac{1}{\overline{x}}.$$

$H_0$ is rejected if the computed value exceeds the upper $\alpha$ point of a $\chi^2_k$ distribution. For reasonable $\alpha$ and $n$ (say $\alpha \approx .05$, $n \geq 10k$) the actual size of the test will be quite near the nominal level $\alpha$ provided that the entries of $X$ are of comparable magnitude (so that the central limit theorem can take effect. See [2] and [3] for a discussion of this point.) In any case, the approximation will be conservative in a moments sense. The discussion in [4] strongly indicates that in those "bad" cases where the central limit theorem fails drastically, the size of the test will tend to err in the conservative direction.

An improvement over the simple $\chi^2_k$ approximation to the distribution of $nR^2$ is suggested by the calculations of section 6:

$$nR^2 \approx c\chi^2_d,$$

where

$$c = 1 - r$$

and

$$d = k\left(\frac{1}{1-r}\right).$$
Here \( r \) is a quantity calculated from the rows \( x_1 \) of \( X \),

\[
r = \frac{\sum \limits_{i=1}^{n} (x_1'X)^{-1}x_1'^2}{k}.
\]

The choice of \( c \) and \( d \) matches the first and second moments of the \( cX_d^2 \) distribution to the corresponding conditional moments of \( nR^2 \) given \( \Phi(X) \).

5. Some "Non-Parametric" Versions of Hotelling's \( T^2 \).

If the random matrix \( X \) has orthant symmetry under the null hypothesis then so will the matrix

\[
\tilde{X} = \left( \text{sign } x_{ij} \right)_{i=1,2,\ldots,n} \quad j=1,2,\ldots,k
\]

obtained by replacing each entry of \( X \) with its sign (or with 0 if \( x_{ij} = 0 \)). The Hotelling's statistic \( \tilde{n}R^2 \) computed from this matrix is an obvious multivariate generalization of the sign test. The heuristic arguments given above for approximating the null distribution of \( nR^2 \) by \( X_k^2 \) have increased plausibility for \( \tilde{n}R^2 \) since the entries of \( \tilde{X} \) are all of the same magnitude, and the central limit theorem can be expected to give good results. Also, of course, the theorem on moments applies as well to \( \tilde{n}R^2 \) as to \( nR^2 \).

Similarly, we can generalize Wilcoxon's signed rank test in a number of different ways. For example, each element of \( X \) can be replaced by the signed rank of its magnitude among all \( nk \) elements (à la Kruskal-Wallis), or, instead, by its signed rank among the \( n \) entries of its column. Either ranking procedure yields a random matrix \( \tilde{X} \) having
orthant symmetry, provided that $X$ has this property. As before, the resulting statistics can be tested against a $\chi^2_k$ null distribution (perhaps more confidently using the second procedure where the entries of $\tilde{X}$ will be more equal in magnitude).

The choice of a "best" version of Hotelling's statistic has not been investigated by the authors. Sometimes, however, the raw value $nR^2$ will be obviously inappropriate. This will be the case, for instance, if egregious outliers are spotted in the data matrix, in which case $nR^2$ can have very poor power characteristics.

In our framework it is possible to look at the data before choosing an appropriate test statistic without violating the $\alpha$-level*, if only properties of $\mathcal{B}(X)$ and not $X$ itself are used to make this choice. For instance, the quantity $r$ of section 4 can be computed from $\mathcal{B}(X)$, and a "non-parametric" test of the type above employed if $r$ is uncomfortably large. For a more extended discussion of the choice of tests in the case $k=1$ the reader is referred to sections 7 and 8 of [4].

6. 

**Proof of the Main Theorem and Related Results.**

In this section, we prove and elaborate the theorem on moments presented in section 3. A more general theorem, of interest outside the context of the present paper, is considered first.

* More precisely, the conditional $\alpha$-level given $\mathcal{B}(X)$. 
Theorem 6.1: Let $U = (U_1, U_2, \ldots, U_n)$ and $V = (V_1, V_2, \ldots, V_n)$ be two random vectors, each of which has independent components distributed symmetrically about 0, and suppose that

\[ \xi(U_i^{2p}) \leq \xi(V_i^{2p}) \; ; \; i=1, \ldots, n; \; p = 1, 2, \ldots . \]  

If $A$ is an $n \times n$ positive semi-definite matrix, then

\[ \xi(UAU')^p \leq \xi(VAV')^p; \; p = 1, 2, \ldots . \]  

Proof: First write $A = BB'$ where $B$ is $n \times n$ and let $b_1, \ldots, b_n$ denote the rows of $B$. Let $W = (W_1, \ldots, W_n)$ be a random vector, where the $W_i$ are independent and each distributed as either $U_i$ or $V_i$.

Then we have

\[ \xi(WBB'W')^p = \xi \left( \left( \sum_{i=1}^{n} W_i b_i \right) \left( \sum_{i=1}^{n} W_i b'_i \right) \right)^p . \]  

Now, suppose that $W_j = U_j$ for some particular $j$. From (6.3), we obtain

\[ \xi(WBB'W')^p = \xi \left( (W_j b_j + \sum_{i \neq j} W_i b_i)(W_j b'_j + \sum_{i \neq j} W_i b'_i) \right)^p = \xi \left( W_j^2 b_j b'_j + 2W_j \sum_{i \neq j} W_i b_i b'_i + Z \right)^p \]

where $Z = \sum_{i \neq j} \sum_{i' \neq j} W_i b_i b'_i b'_i b'_i$. Using the trinomial expansion on the last line of (6.4), it follows that
(6.5) \[ \mathbf{\xi}^{(WBB'W')}^P = \]
\[ \sum_{\alpha_1, \alpha_2, \alpha_3} \left( \begin{array}{c} \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_1 + \alpha_2 + \alpha_3 = p \end{array} \right) \mathbf{U}_j^{2\alpha_1 + \alpha_2} \mathbf{b}_j^{\alpha_2} \mathbf{b}_j^{\alpha_3} \left( 2 \sum_{i \neq j} \mathbf{W}_i \mathbf{b}_i^{\alpha_1} \mathbf{b}_j^{\alpha_2} \mathbf{Z}_j^{\alpha_3} \right) \]
\[ \leq \mathbf{\xi} \sum_{\alpha_1, \alpha_2, \alpha_3} \left( \begin{array}{c} \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_1 + \alpha_2 + \alpha_3 = p \end{array} \right) \mathbf{V}_j^{2\alpha_1 + \alpha_2} \mathbf{b}_j^{\alpha_2} \mathbf{b}_j^{\alpha_3} \left( 2 \sum_{i \neq j} \mathbf{W}_i \mathbf{b}_i^{\alpha_1} \mathbf{b}_j^{\alpha_2} \mathbf{Z}_j^{\alpha_3} \right) \]
\[ = \mathbf{\xi}^{(\tilde{W}BB'\tilde{W}')}^P \]

where \( \tilde{W} \) is the vector \( W \) with the distribution of \( W_j \) changed from that of \( U_j \) to that of \( V_j \). The inequality in (6.5) follows from assumption (6.1), symmetry about 0, and the independence of \( (W_1, \ldots, W_n) \). (Note that only those cases where \( \alpha_2 \) is even contribute to the expectation, and in these cases the coefficient of the \( W_j \) term, \( (b_j b_j')^{\alpha_2} (2 \sum_{i \neq j} W_i b_i b_j')^{\alpha_3} \) is non-negative with probability 1.)

Now, starting with \( W = U \), successively replacing \( W_j \) by \( V_j \), and applying the above argument, (6.5) implies that (6.2) holds. This completes the proof.

As an application of this result, consider a random matrix \( X : n \times k \) which has orthant symmetry and rank \( k \) with probability one, \( k \leq n \).

Then the distribution of \( \mathbf{\xi}(X'X)^{-1}X'\tilde{\mathbf{e}} \) can be represented as a mixture of distributions of the form \( \Delta C'C' \) where \( \tilde{\mathbf{e}} = (1,1,\ldots,1) : 1 \times n \), \( \Delta \) is a vector of independent symmetric \( \mathbf{+1} \) random variables, and \( C : n \times k \) is such that \( C'C = I \).

**Remark:** The assumption that \( X \) has full rank with probability one is for convenience. Since the argument given below is conditional on \( X \),
the general case when rank of $X$ is $j$ with probability $p_j$, $0 \leq j \leq k$, can be treated by the same methods as the case of full rank, yielding the bound $\mathcal{E}(\chi_j^2)^p$ which is less than $\mathcal{E}(\chi_k^2)^p$ if $j < k$.

From Theorem 6.1, we have

**Corollary 6.1:** For each $p = 1, 2, \ldots$,

\begin{equation}
\mathcal{E}(\tilde{e}X(X'X)^{-1}X'e)^p \leq \mathcal{E}(\chi_k^2)^p
\end{equation}

where $\chi_k^2$ is a chi-square random variable with $k$ degrees of freedom.

**Proof:** It is sufficient to show that

\begin{equation}
\mathcal{E}(\Delta C'C')^p \leq \mathcal{E}(\chi_k^2)^p
\end{equation}

for each $C$: $n \times k$ such that $C'C = I_k$. Application of Theorem 6.1 with $U = \Delta$ and $V_i \sim \mathcal{N}(0, 1)$ yields

\begin{equation}
\mathcal{E}(\Delta C'C')^p \leq \mathcal{E}(VCC'V')^p.
\end{equation}

Since $C'C = I_k$, we see that $VCC'V'$ has the distribution of $\chi_k^2$ and the result follows.

A question which arises in connection with the above result is that of maximizing (over $C$) the left hand side of (6.7) for a fixed $n$ and $k$. For $k=1$, it has been shown in [4] that $\mathcal{E}(\Delta C'C')^p$ is maximized by $C = \tilde{e}'/\sqrt{n}$.

We now discuss the case for general $k$ and $n$ for $p = 2$. Let $C$: $n \times k$ have rows $\xi_1, \ldots, \xi_n$ such that $C'C = I_k$. A straightforward calculation shows that
\[ (6.9) \quad \xi(x_k^2)^2 - \xi(\Delta C'\Delta')^2 = 2 \sum_{1}^{n} \|\xi_1\|^4. \]

The following proposition gives upper and lower bounds for (6.9).

**Proposition 6.1:** For any matrix \( C : nxk \), with rows \( \xi_1', \ldots, \xi_n' \), such that \( C'C = I_k \), we have

\[ (6.10) \quad \frac{k^2}{n} \leq \sum_{1}^{n} \|\xi_1\|^4 \leq k. \]

**Proof:** We first establish the upper bound in (6.10). Recall that if \( A_1 \) and \( A_2 \) are \( nxn \) real matrices, then the bilinear function

\[ < A_1A_2 > = \text{tr} A_1A_2' \]

defines an inner product on the space of all \( nxn \) real matrices. Let \( \pi \) be a linear transformation of this space onto itself defined as follows: \( \pi(A) \) takes \( A = \{ a_{ij} \} \) into the diagonal matrix with diagonal elements \( a_{11} \). It follows directly from the definition that \( \pi \) is an orthogonal projection with respect to the above inner product.

Now, it is clear that

\[ (6.11) \quad < \pi(\Delta C'), \pi(\Delta C') > = \sum_{1}^{n} \|\xi_1\|^4. \]

Since the norm of a projected vector is no larger than the norm of the vector, we have

\[ (6.12) \quad \sum_{1}^{n} \|\xi_i\|^4 \leq < \Delta C', \Delta C' > = \text{tr} C'C \| CC'C \| = \text{tr} I_k = k. \]

The upper bound is established.
To obtain the lower bound, let \( z_{i,j} = c_{i,j}^2 \) and note that

\[
\sum_{i=1}^{n} \sum_{j=1}^{k} |s_i|^4 = \sum_{i=1}^{n} \sum_{j=1}^{k} \sum_{j'=1}^{k} c_{i,j}^2 c_{i,j'}^2 = \sum_{j=1}^{k} \sum_{j'=1}^{k} \sum_{i=1}^{n} z_{i,j} z_{i,j'} = \sum_{j=1}^{k} \sum_{j'=1}^{k} (z(j), z(j'))
\]

\[
= \| \sum_{j=1}^{k} z(j) \|^2
\]

where \( z(j) = (z_{1,j}, \ldots, z_{n,j}) \). Since \( C'C = I_k \), \( z(j)e_i = 1 \) for \( j = 1, \ldots, k \). Then, from the Cauchy-Schwartz Inequality, we have

\[
k^2 = \| \sum_{j=1}^{k} z(j) \|^2 \leq \| \sum_{j=1}^{k} z(j) \| \cdot \| \sum_{j=1}^{k} z(j) \| = n\| \sum_{j=1}^{k} z(j) \|^2
\]

since \( \| e \|^2 = n \). This completes the proof.

Note that the upper bound in (6.10) is achieved for all \( k \) and \( n \), \( k \leq n \), by choosing the first \( k \) rows of \( C \) to be the standard orthonormal basis in \( R^k \) and the remaining rows \( 0 \). Also, it follows from the derivation of the lower bound that we have equality for the lower bound if and only if

\[
( \sum_{j=1}^{k} z(j) = \frac{k}{n} e_i )
\]

that is, if and only if the norm squared of each row of \( C \) is \( k/n \).
For each \( k \) and \( n \), \( 1 \leq k \leq n \), let \( M(k,n) \) be the set of matrices \( C: n \times k \) such that \( C'C = I_k \) and each row of \( C \) has norm squared \( k/n \). Also, let \( \mathcal{A} = \{ (k,n) | M(k,n) \neq \emptyset \} \), where \( \emptyset \) denotes the empty set.
Proposition 6.2: The set $M(k,n)$ is non-empty for each $(k,n)$, $1 \leq k \leq n$.

Proof: We first list four properties of the set $\mathcal{A}$.

(i) $(1,n) \in \mathcal{A}$ since $e'/\sqrt{n} \in M(1,n)$.

(ii) $(k,m) \in \mathcal{A}$ for $k=1, \ldots$ and $m=1, \ldots$. To show this, let $\Gamma$ be $k \times k$ and orthogonal and let $C' = \frac{1}{\sqrt{m}} (\Gamma', \Gamma', \ldots, \Gamma')$ be $k \times mk$. It is easy to show that $C \in M(k,m)$.

(iii) If $(k,n) \in \mathcal{A}$, then $(n-k,n) \in \mathcal{A}$. For $(k,n) \in \mathcal{A}$, let $C_1 \in M(k,n)$ and complete $C_1$ to $\Gamma$, an $n \times n$ orthogonal matrix. Then $C_2 : n \times (n-k)$ consisting of the last $(n-k)$ columns of $\Gamma$ is in $M(n-k,n)$.

(iv) If $(k,n) \in \mathcal{A}$, then $(k,n+k) \in \mathcal{A}$. For this assertion, let $C \in M(k,n)$ and define $C_1 : (n+k) \times k$ by $C_1 = \frac{1}{\beta} \left( \begin{array}{c} C \\ \beta I_k \end{array} \right)$ where $\beta = \sqrt{k/n}$. It is easy to show that $C_1 \in M(k,n+k)$ so that $(k,n+k) \in \mathcal{A}$.

Now, let $(k,n)$ be such that $1 \leq k \leq n$ (if $k=n$, then clearly $(n,n) \in \mathcal{A}$). Write $n = dk + \ell$ where $d$ is an integer and $0 \leq \ell < k$. If $\ell = 0$, then by (ii), $(k,n) \in \mathcal{A}$. For the case $1 \leq \ell < k$, we must show that $(k,dk + \ell) \in \mathcal{A}$. By (iv), it is sufficient to show that $(k,k+\ell) \in \mathcal{A}$. Then by (iii), it is sufficient to show that $(\ell,k + \ell) \in \mathcal{A}$. However, applying (iv) again, it is sufficient to prove that $(\ell,k) \in \mathcal{A}$.

If $\ell = 1$, then application of (i) yields the result. If $\ell > 1$, we simply apply the above argument again to $(\ell,k)$. Continuing in this manner at most $k$ times, we either apply (ii) at some stage to obtain the result or we continue to eventually apply (i) and obtain the desired conclusion. This completes the proof.
From Proposition (6.2), we conclude that the lower bound in (6.10) is achieved for all values of \( k \) and \( n, \ k \leq n \). Now, suppose \( n \) is such that there exists an \( n \times n \) orthogonal matrix \( H = \{h_{ij}\} \) where 
\[ h_{ij} = \pm \frac{1}{\sqrt{n}}. \]
Such a matrix is called a Hadamard Matrix. In this special case, note that \( C: n \times k \) consisting of the first \( k \) columns of \( H \) attains the lower bound in (6.10) for each \( k \). Sufficient conditions on \( n \) for the existence of \( H \) are given in Paley (1933).

For \( p = 3 \), a tedious calculation also establishes

\[
(6.16) \quad \mathcal{E}(\chi_k^2)^3 - \mathcal{E}(\Delta C \Delta')^3 = \\
(20+6k) \sum_1^n \| \xi_i \|^4 - 16 \sum_1^n \| \xi_i \|^6.
\]

The authors do not have tight bounds for (6.16). However, substitution of the two extreme cases in (6.10) yield \( 6k^2+4k \) for the upper bound case and \( \frac{k^2}{n} \left[ 20+6k - 16 \frac{k}{n} \right] \) for the lower bound case for (6.16).
REFERENCES


