ON THE PROBLEM OF TESTING LOCATION IN MULTIVARIATE POPULATIONS FOR RESTRICTED ALTERNATIVES[1]

By Peter E. Nüesch[2]

University of California, Berkeley

0. Introduction. In a comprehensive paper Bartholomew [2] treats the problem of testing equality of means in normal populations versus ordered and partially ordered alternatives. The application of the likelihood ratio (LR) principle to derive the test statistic leads to minimize a convex quadratic form subject to linear constraints. The author uses a theorem by van Eeden [7] to solve this problem but remarks that this could also be done by means of quadratic programming methods. This remark is the starting point to our investigation.

It is easy to see that the setup of Bartholomew is a special case of the following problem: Let $\mathbf{X}$ be a p-dimensional random vector, multivariate normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The null hypothesis H, that the center of the distribution lies at a point $\mathbf{0}$, is to be tested versus the alternative K that it lies in the union of some regions R_i, where each R_i is the intersection of p halfspaces through $\mathbf{0}$. (Without loss of generality we choose $\mathbf{0}$ as the origin of the coordinate system.) We may restrict ourselves to the alternative that $\boldsymbol{\mu}$ lies in the positive orthant $K_0 : \boldsymbol{\mu} \geqq \mathbf{0}$ with at least one strict inequality ($\boldsymbol{\mu} > \mathbf{0}$ denotes that all components of the vector are positive) since the methods developed extend immediately if the orthant is replaced by the union of arbitrarily many R_i. The extension has been carried out in more details in Nüesch [5].

Bartholomew's results follow from ours by specializing to particular covariance matrices. (e.g. the covariance matrix corresponding to completely ordered parameters is a so-called type 2 matrix with entries $\rho_{ij} = 0$ for $|i - j| > 2$.)

In Section 1 quadratic programming methods are used to minimize a general convex quadratic form $f(\mathbf{y})$ subject to the constraints $\mathbf{y} \geqq \mathbf{0}$. In Sections 2 and 3 the test statistics and their distributions are derived for the cases of known and unknown covariance matrix, respectively.[3]

1. Solutions of the reduced quadratic programming problem. Let $\mathbf{C}$ be a positive definite $p \times p$ matrix ($\mathbf{C} > \mathbf{0}$), $\mathbf{q}$ a $1 \times p$ matrix of constants. The function

$$f(\mathbf{y}) = 2\mathbf{q}\mathbf{y} + \mathbf{y}'\mathbf{C}\mathbf{y} \tag{1.1}$$

Received 22 November 1963; revised 19 April 1965.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

[2] Now at Kantonale Oberrealschule, Zurich.

[3] Section 2 overlaps with results by Kudô [4], which were published after this paper was submitted. Kudô uses an intuitive geometrical argument to obtain the maximum likelihood estimates. This argument however does not carry over to the case of unknown covariance matrix as does our approach.

is a convex quadratic form in $\mathbf{y}$. The problem is to minimize $f(\mathbf{y})$ subject to $\mathbf{y} \geqq \mathbf{0}$. (In this reduced quadratic programming problem the additional condition $\mathbf{Ay} = \mathbf{b}$ is missing. The absence of this latter constraint makes it possible to arrive at the solution without the usual stepwise procedure for solving quadratic programming problems.) We call a vector $\mathbf{y}$ minimizing (1.1) *optimal.*

Eisenberg [3] proved that a necessary and sufficient condition for $\hat{\mathbf{y}}$ to be optimal is

$$(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} \leqq (\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\mathbf{y} \tag{1.2}$$

whenever $\mathbf{y} \geqq \mathbf{0}$. Inequality (1.2) enables us to prove the following

LEMMA 1.1. $\hat{\mathbf{y}}$ *is optimal if and only if*

$$q_j + \hat{\mathbf{y}}'\mathbf{C}^j \geqq 0 \tag{1.3}$$

and

$$q_j + \hat{\mathbf{y}}'\mathbf{C}^j > 0 \quad \textit{implies} \quad \hat{y}_j = 0 \quad \textit{for all} \quad j, \tag{1.4}$$

where q_j *denotes the* jth *component of* q, $\mathbf{C}^j$ *the* jth *column of* $\mathbf{C}$.

Using identification (2.5), Lemma 1.1 becomes Theorem 2.1 of Kudô [4].

PROOF. Suppose that for some j $(q_j + \hat{\mathbf{y}}'\mathbf{C}^j) < 0$. Let $y_j = a$, $y_i = 0$ for $i \neq j$. Then

$$(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\mathbf{y} = (q_j + \hat{\mathbf{y}}'\mathbf{C}^j)a. \tag{1.5}$$

Let a tend to $+\infty$. The right side of (1.5) tends to $-\infty$ and the inequality (1.2) becomes impossible. (1.3) is proved. Suppose now $(q_j + \hat{\mathbf{y}}'\mathbf{C}^j) > 0$. We want to show $\hat{y}_j = 0$. Assume this is not the case, i.e., $\hat{y}_j > 0$. Let $y_i = 0$ for $i \neq j$ and $y_j = \hat{y}_j/2$. This gives

$$(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} \geqq (q_j + \hat{\mathbf{y}}'\mathbf{C}^j)\hat{y}_j > (q_j + \hat{\mathbf{y}}'\mathbf{C}^j)y_j = (\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\mathbf{y},$$

which contradicts (1.2). (1.4) is proved.

To prove the converse, note that (1.3) and (1.4) imply

$$(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} = \mathbf{0}, \tag{1.6}$$

since $(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} = \sum_{j=1}^{p} (q_j + \hat{\mathbf{y}}'\mathbf{C}^j)\hat{y}_j$ and for each j either one of the two factors is zero. (1.3) implies $\mathbf{0} \leqq (\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\mathbf{y}$ for all $\mathbf{y} \geqq \mathbf{0}$. Using (1.6) we get $(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} \leqq (\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\mathbf{y}$ which is exactly inequality (1.2). Hence $\hat{\mathbf{y}}$ satisfying (1.3) and (1.4) is optimal.

Using (1.6) we get the minimal value of f as

$$f(\hat{\mathbf{y}}) = 2(\mathbf{q} + \hat{\mathbf{y}}'\mathbf{C})\hat{\mathbf{y}} - \hat{\mathbf{y}}'\mathbf{C}\hat{\mathbf{y}} = -\hat{\mathbf{y}}'\mathbf{C}\hat{\mathbf{y}}. \tag{1.7}$$

2. The test in the case of known covariance matrix. Let $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be a sample of size N from $\mathbf{X}$, where $N > p$ to avoid degeneracies. The likelihood function is

$$L = [(2\pi)^{\frac{1}{2}pN} |\mathbf{\Sigma}|^{\frac{1}{2}N}]^{-1} \exp [-\tfrac{1}{2} \sum (\mathbf{x}_\alpha - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu})], \tag{2.1}$$

where summation, unless otherwise specified, runs from 1 to N. Denote the maximum of L over ω and Ω by $L(\hat{\omega})$ and $L(\hat{\Omega})$, respectively. (ω, Ω are the parts of the parameter space specified by the hypothesis and the alternative, respectively.) Since ω consists of the single point $\{\boldsymbol{\mu} = \mathbf{0}\}$, the maximum likelihood estimate $\hat{\boldsymbol{\mu}}_\omega$ is equal to $\mathbf{0}$, and since $\boldsymbol{\Sigma}$ is known, we have

$$L(\omega) = L(\hat{\omega}) = k\cdot \exp\left[-\tfrac{1}{2}\sum \mathbf{x}_\alpha' \boldsymbol{\Sigma}^{-1}\mathbf{x}_\alpha\right]. \tag{2.2}$$

Maximizing

$$L(\Omega) = k\cdot \exp\left[-\tfrac{1}{2}\sum (\mathbf{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_\alpha - \boldsymbol{\mu})\right] \tag{2.3}$$

over Ω means finding that $\boldsymbol{\mu} \geqq \mathbf{0}$ which minimizes the exponent or equivalently

$$f(\boldsymbol{\mu}) = -2\bar{\mathbf{x}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{2.4}$$

over Ω. Denote it by $\hat{\boldsymbol{\mu}}$. (We omit the subscript Ω, since in the following we are exclusively concerned with $\hat{\boldsymbol{\mu}}_\Omega$.) By setting

$$\boldsymbol{\Sigma}^{-1} = \mathbf{C}, \qquad -\bar{\mathbf{x}}'\boldsymbol{\Sigma}^{-1} = \mathbf{q}, \tag{2.5}$$

(2.4) becomes (1.1). Since $\boldsymbol{\Sigma}$ and thus $\boldsymbol{\Sigma}^{-1}$ is assumed to be positive definite, the results of Section 1 are applicable. From (1.7) we obtain the minimal value of $f(\boldsymbol{\mu})$ as $f(\hat{\boldsymbol{\mu}}) = -\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$ and $L(\Omega)$ is therefore maximized by

$$L(\hat{\Omega}) = k\cdot \exp\left[-\tfrac{1}{2}\left(\sum \mathbf{x}_\alpha'\boldsymbol{\Sigma}^{-1}\mathbf{x}_\alpha - N\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}\right)\right]. \tag{2.6}$$

The LR criterion rejects for $\lambda = L(\hat{\omega})/L(\hat{\Omega}) \leqq \lambda_0$. This proves

THEOREM 2.1. *The* LR *test for testing H versus K_0 rejects when*

$$N\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} \geqq c^2 \tag{2.7}$$

where $\hat{\boldsymbol{\mu}}$ minimizes (2.4) *subject to the constraint $\boldsymbol{\mu} \geqq \mathbf{0}$.*

For the actual computation of the maximum likelihood estimate when $\bar{\mathbf{x}}$ is given the reader is referred to Kudô ([4], Section 2) or Nüesch ([5], Section 3). Geometrically speaking the maximum likelihood estimate is the projection of the vector $\bar{\mathbf{x}}$ along regression planes onto the positive orthant of the sample space. (If one uses a linear transformation of $\mathbf{X}$ to an uncorrelated $\mathbf{Y}$, which exists for a full rank covariance matrix, the projection is orthogonal onto a polyhedral half cone, the affine image of the positive orthant.) The first author also gives an example of the computation at the end of Section 2.

$\hat{\boldsymbol{\mu}}$ is a vector whose components are either positive or zero. This leads to a partition of the sample space into 2^p disjoint regions. Denote by $\mathfrak{X}_k$ any of the $\binom{p}{k}$ regions of the sample space with exactly k of the $\hat{\mu}_j$'s positive. Since through relabeling of the variables one always can obtain that the k positive $\hat{\mu}_j$'s are the last k, we assume that

$$\mathfrak{X}_k = \{\bar{\mathbf{x}};\ \{\hat{\boldsymbol{\mu}}_{(1)} = \mathbf{0}\} \text{ and } \{\hat{\boldsymbol{\mu}}_{(2)} > \mathbf{0}\}\}, \tag{2.8}$$

where $\hat{\boldsymbol{\mu}}' = (\hat{\boldsymbol{\mu}}'_{(1)}\,\hat{\boldsymbol{\mu}}'_{(2)})$. Let $\mathbf{C}$ be partitioned accordingly:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}.$$

(1.3), (1.4) and the identification (2.5) enable us to define $\mathfrak{X}_k$ in terms of $\bar{\mathbf{x}}_{(1)}$ and $\bar{\mathbf{x}}_{(2)}$. $\hat{\mathbf{u}}_{(1)} = \mathbf{0}$ and $\hat{\mathbf{u}}_{(2)} > \mathbf{0}$ imply

$$-\mathbf{C}_{11}\bar{\mathbf{x}}_{(1)} + \mathbf{C}_{12}(\hat{\mathbf{u}}_{(2)} - \bar{\mathbf{x}}_{(2)}) > \mathbf{0}, \tag{2.9}$$

$$-\mathbf{C}_{21}\bar{\mathbf{x}}_{(1)} + \mathbf{C}_{22}(\hat{\mathbf{u}}_{(2)} - \bar{\mathbf{x}}_{(2)}) = \mathbf{0}. \tag{2.10}$$

Solving (2.10) we get

$$\hat{\mathbf{u}}_{(2)} = \bar{\mathbf{x}}_{(2)} + \mathbf{C}_{22}^{-1}\mathbf{C}_{21}\bar{\mathbf{x}}_{(1)} = \bar{\mathbf{x}}_{(2)} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\bar{\mathbf{x}}_{(1)} . \tag{2.11}$$

Using (2.11), (2.9) then reads

$$-(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21})\bar{\mathbf{x}}_{(1)} = -\mathbf{\Sigma}_{11}^{-1}\bar{\mathbf{x}}_{(1)} > \mathbf{0} \tag{2.12}$$

and the definition (2.8) becomes

$$\mathfrak{X}_k = \{\bar{\mathbf{x}};\ \{\mathbf{\Sigma}_{11}^{-1}\bar{\mathbf{x}}_{(1)} < \mathbf{0}\} \cap \{\bar{\mathbf{x}}_{(2)} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\bar{\mathbf{x}}_{(1)} > \mathbf{0}\}\}. \tag{2.13}$$

We now derive the distribution of the test (2.7).

THEOREM 2.2. *The test statistic* $N\hat{\mathbf{u}}'\mathbf{\Sigma}^{-1}\hat{\mathbf{u}} \geqq c^2$ *is under* H *distributed as* $\sum_{k=1}^{p} w(p, k)P(\chi_k^2 \geqq c^2)$ *where the weights* $w(p, k)$ *are the probability content of the union of all* $\mathfrak{X}_k$'s *for a fixed* k.

PROOF. Denote the rejection region by S^c.

$$P(S^c) = \sum_{k=1}^{p} \sum_{\mathfrak{X}_k} P(S^c \cap \mathfrak{X}_k) = \sum_{k=1}^{p} \sum_{\mathfrak{X}_k}\{P(\mathfrak{X}_k)P(S^c \mid \mathfrak{X}_k)\}.$$

The summation starts at 1 since $S^c \cap \mathfrak{X}_0 = \varnothing$.

$$P(S^c \mid \mathfrak{X}_k) = P[(N\hat{\mathbf{u}}'\mathbf{\Sigma}^{-1}\hat{\mathbf{u}} \geqq c^2) \mid \{\hat{\mathbf{u}}_{(1)} = \mathbf{0}\} \cap \{\hat{\mathbf{u}}_{(2)} > \mathbf{0}\}]. \tag{2.14}$$

Since $N^{\frac{1}{2}}\hat{\mathbf{u}}_{(2)}$ is, under H, normally distributed with expectation 0 and covariance matrix $\mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{12} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$, it follows that $\hat{\mathbf{u}}_{(2)}'\mathbf{\Sigma}_{22\cdot1}^{-1}\hat{\mathbf{u}}_{(2)}$ is χ_k^2-distributed. In addition $\{\hat{\mathbf{u}}_{(2)} > \mathbf{0}\}$ and $\{\hat{\mathbf{u}}_{(2)}'\mathbf{\Sigma}_{22.1}^{-1}\hat{\mathbf{u}}_{(2)} \geqq c^2\}$ are independent. Therefore (2.14) becomes

$$P(S^c \mid \mathfrak{X}_k) = P[N\hat{\mathbf{u}}_{(2)}'\mathbf{\Sigma}_{22.1}^{-1}\hat{\mathbf{u}}_{(2)} \geqq c^2] = P(\chi_k^2 \geqq c^2),$$

which does not depend upon the particular $\mathfrak{X}_k$. We then have

$$P(S^c) = \sum_{k=1}^{p} \{\sum_{\mathfrak{X}_k} P(\mathfrak{X}_k)\}P(S^c \mid \mathfrak{X}_k) = \sum_{k=1}^{p} w(p, k)P(\chi_k^2 \geqq c^2)$$

which proves the theorem.

3. The case of unknown covariance matrix. In the likelihood function given by (2.1) $\mathbf{\Sigma}$ is now to be considered as variable. As before, let $\mathbf{C} = \mathbf{\Sigma}^{-1}$. From (2.2) we get

$$\begin{aligned}\log L(\omega) &= -\tfrac{1}{2}pN \log (2\pi) + \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \sum \mathbf{x}_\alpha'\mathbf{C}\mathbf{x}_\alpha \\ &= -\tfrac{1}{2}pN \log (2\pi) + \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \operatorname{tr} (\mathbf{C}\mathbf{D}_1),\end{aligned}$$

where $\mathbf{D}_1 = \sum \mathbf{x}_\alpha\mathbf{x}_\alpha'$ is positive definite. Anderson ([1], Lemma 3.2.2) gives the maximum immediately as

$$\log L(\hat{\omega}) = -\tfrac{1}{2}pN \log (2\pi) + \tfrac{1}{2}Np \log N - \tfrac{1}{2}N \log |\mathbf{D}_1| - \tfrac{1}{2}Np. \tag{3.1}$$

Now we want to maximize $\log L(\Omega)$ over the set of all positive definite $\mathbf{C}$, $\{\mathbf{C} > \mathbf{0}\}$, and $\{\boldsymbol{\mu} \geqq \mathbf{0}\}$. From (2.3) we have

$$\log L(\Omega) = -\tfrac{1}{2}pN \log (2\pi) + \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \sum (\mathbf{x}_\alpha - \boldsymbol{\mu})'\mathbf{C}(\mathbf{x}_\alpha - \boldsymbol{\mu}).$$

Let

$$f(\boldsymbol{\mu}, \mathbf{C}) = \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \sum (\mathbf{x}_\alpha - \boldsymbol{\mu})'\mathbf{C}(\mathbf{x}_\alpha - \boldsymbol{\mu}),$$

but

$$\max_{\{\boldsymbol{\mu} \geqq \mathbf{0}\}, \{\mathbf{C} > \mathbf{0}\}} f(\boldsymbol{\mu}, \mathbf{C}) = \max_{\{\mathbf{C} > \mathbf{0}\}} \{\max_{\{\boldsymbol{\mu} \geqq \mathbf{0}\}} f(\boldsymbol{\mu}, \mathbf{C})\}.$$

Now

$$f(\mathbf{C}) = \max_{\{\boldsymbol{\mu} \geqq \mathbf{0}\}} f(\boldsymbol{\mu}, \mathbf{C}) = \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \min_{\{\boldsymbol{\mu} \geqq \mathbf{0}\}} [\textstyle\sum (\mathbf{x}_\alpha - \boldsymbol{\mu})'\mathbf{C}(\mathbf{x}_\alpha - \boldsymbol{\mu})].$$

This is the same problem that was solved in Section 2. We get therefore from (2.6)

$$\begin{aligned} f(\mathbf{C}) &= \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2}[\textstyle\sum \mathbf{x}_\alpha'\mathbf{C}\mathbf{x}_\alpha - N\hat{\boldsymbol{\mu}}'\mathbf{C}\hat{\boldsymbol{\mu}}] \\ &= \tfrac{1}{2}N \log |\mathbf{C}| - \tfrac{1}{2} \operatorname{tr} (\mathbf{C}\mathbf{D}_2), \end{aligned}$$

where

$$\mathbf{D}_2 = \sum \mathbf{x}_\alpha \mathbf{x}_\alpha' - N\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}'. \tag{3.2}$$

Note that $\mathbf{D}_2$ is also positive definite.

We may rewrite $\mathbf{D}_2$ as

$$\begin{aligned} \mathbf{D}_2 &= \sum (\mathbf{x}_\alpha - \hat{\boldsymbol{\mu}})(\mathbf{x}_\alpha - \hat{\boldsymbol{\mu}})' = \sum (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})' \\ &\quad + N(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})' \\ &= \mathbf{A} + N(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})', \end{aligned} \tag{3.3}$$

where $\mathbf{A}$ is the sample covariance matrix. Using again Anderson ([1], Lemma 3.2.2) we get

$$\log L(\hat{\Omega}) = \max_{\{\mathbf{C} > \mathbf{0}\}} f(\mathbf{C}) = \tfrac{1}{2}pN \log N - \tfrac{1}{2}N \log |\mathbf{D}_2| - \tfrac{1}{2}pN. \tag{3.4}$$

From (3.1), (3.2) and (3.4) we obtain

$$\lambda^{2/N} = \{L(\hat{\omega})/L(\hat{\Omega})\}^{2/N} = |\mathbf{D}_2|/|\mathbf{D}_1| = [1 + N\hat{\boldsymbol{\mu}}'\mathbf{D}_2^{-1}\hat{\boldsymbol{\mu}}]^{-1}.$$

This proves

THEOREM 3.1. *The* LR *test for testing H versus K_0 in the case of unknown covariance matrix $\boldsymbol{\Sigma}$ rejects if*

$$N\hat{\boldsymbol{\mu}}'\mathbf{D}_2^{-1}\hat{\boldsymbol{\mu}} \geqq c^2 \tag{3.5}$$

where $\mathbf{D}_2$ is given by (3.3).

We now derive the distribution of the statistic $N\hat{\boldsymbol{\mu}}'\mathbf{D}_2^{-1}\hat{\boldsymbol{\mu}}$.

$N^{\frac{1}{2}}(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})$ is normally distributed with expectation zero and covariance matrix

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{pmatrix}.$$

The random variable $\mathbf{A}$ is Wishart distributed with parameters p, $\mathbf{\Sigma}$, $N - 1$ and is independent of $N^{\frac{1}{2}}(\bar{\mathbf{x}} - \hat{\mathbf{\mu}})$. So $\mathbf{D}_2$ is the sum of two independent Wishart matrices with different parameters, which is not a Wishart matrix, and the distribution of the test statistic is not of known form. However, in Theorem 3.2 we give an upper and lower stochastic bound with familiar distributions.

THEOREM 3.2. *Over $\mathfrak{X}_k$ the statistic $N\hat{\mathbf{\mu}}'\mathbf{D}_2^{-1}\hat{\mathbf{\mu}}$ is stochastically larger than a χ_k^2/χ_{N-p+1}^2-variable and stochastically smaller than a χ_k^2/χ_{N-p}^2-variable.*

PROOF. For a positive definite matrix $\mathbf{G}$ and a matrix $\mathbf{F}$ of rank 1 ($\mathbf{F} = \mathbf{zy}'$) the following expansion holds:

$$(\mathbf{G} + \mathbf{F})^{-1} = \mathbf{G}^{-1} - (1 + \mathbf{y}'\mathbf{G}\mathbf{z})^{-1}(\mathbf{G}^{-1}\mathbf{z})(\mathbf{y}\mathbf{G}^{-1})$$

(see, e.g., Roy-Sarhan [6]). If $\mathbf{G}$ and $\mathbf{F}$ are symmetric ($\mathbf{F} = \mathbf{yy}'$) the matrix is positive definite and the inequality

$$\mathbf{x}'\mathbf{G}^{-1}\mathbf{x} \geqq \mathbf{x}'(\mathbf{G} + \mathbf{F})^{-1}\mathbf{x}, \qquad \mathbf{x} \neq \mathbf{0}, \tag{3.6}$$

holds.

Applying (3.6) twice, the first time with $\mathbf{G} = \mathbf{A}$, $\mathbf{F} = N(\bar{\mathbf{x}} - \hat{\mathbf{\mu}})(\bar{\mathbf{x}} - \hat{\mathbf{\mu}})'$, the second time with $\mathbf{G} = \mathbf{D}_2$, $\mathbf{F} = N\hat{\mathbf{\mu}}\hat{\mathbf{\mu}}'$ yields the stochastic ordering

$$N\hat{\mathbf{\mu}}'\mathbf{A}^{-1}\hat{\mathbf{\mu}} \geqq N\hat{\mathbf{\mu}}'\mathbf{D}_2^{-1}\hat{\mathbf{\mu}} \geqq N\hat{\mathbf{\mu}}'\mathbf{D}_1^{-1}\hat{\mathbf{\mu}}. \tag{3.7}$$

The random variables $\mathbf{A}$ and $\mathbf{D}_1$ have Wishart distributions with parameters p, $\mathbf{\Sigma}$, $N - 1$ and p, $\mathbf{\Sigma}$, N, -respectively. Conditioned on $\hat{\mathbf{\mu}}_{(2)} = 0$ the distributions of $\mathbf{A}_{22.1}$ and $\mathbf{D}_{1_{22.1}}$ are again Wishart distributions with parameters k, $\mathbf{\Sigma}_{22.1}$, $N - 1 - p + k$ and k, $\mathbf{\Sigma}_{22.1}$, $N - p + k$, respectively. The distributions of the variables of the first and third term in (3.7) now follow from [1], Theorem 5.2.2.

In virtue of this theorem and the intractability of the exact distribution of the test statistic, we now modify test (3.5) slightly and

(3.8) reject H if $N\hat{\mathbf{\mu}}'\mathbf{A}^{-1}\hat{\mathbf{\mu}} \geqq c^2$, where $\mathbf{A}$ is the sample covariance matrix.

Since $\bar{\mathbf{x}}$ and $\mathbf{A}$ are independent random variables $\hat{\mathbf{\mu}}_{(2)}$ and $\mathbf{A}_{22.1}$ are also independent and the distribution over $\mathfrak{X}_k$ is therefore the ratio of two independent χ^2-variables.

THEOREM 3.3. *The test statistic $N\hat{\mathbf{\mu}}'\mathbf{A}^{-1}\hat{\mathbf{\mu}} \geqq c^2$ is distributed under H as*

$$\textstyle\sum_{k=1}^{p} w(p, k)P[B(k/2, (N-p)/2) \geqq c^2],$$

where the weights $w(p, k)$ are the same as in Theorem 2.2, and $B(\nu_1, \nu_2)$ is a beta distributed random variable with ν_1, ν_2 degrees of freedom.

PROOF. The modified test statistic (3.8) depends on $\hat{\mathbf{\mu}}$ in the same way as in the case of known covariance matrix. The proof is the same as the one of Theorem 2.2 when $\mathbf{\Sigma}$ is replaced by $\mathbf{A}$. The arguments used there carry over to this case.

4. Remarks. In [5] the weights $w(p, k)$ are computed for dimensions $p = 2, 3$ and for certain correlation structures also for general p. Tables for the cut-off-points c^2 are given for $p = 2$ and the alternatives $K_0 : \mathbf{\mu} \geqq \mathbf{0}$ and $K_0' : \mathbf{\mu} \geqq \mathbf{0}$ or

$\mathbf{\mu} \leq \mathbf{0}$. Furthermore a comparison between the power of the test (2.7) and the usual χ^2-test for unrestricted alternatives is included.

5. Acknowledgments. I wish to thank Professor E. L. Lehmann for recommending the problem and for his guidance throughout the work.

REFERENCES

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

[2] BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B* **23** 239–281.

[3] EISENBERG, E. (1963). A digression on quadratic programming. Mimeographed notes, Department of Industrial Engineering, University of California, Berkeley.

[4] KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50** 403–418.

[5] NÜESCH, P. E. (1964). Multivariate tests of location for restricted alternatives. Doctoral dissertation, Swiss Federal Institute of Technology, Zurich.

[6] ROY, S. N. and SARHAN, A. E. (1956). On inverting a class of patterned matrices. *Biometrika* **43** 227–231.

[7] VAN EEDEN, C. (1958). Testing and estimating ordered parameters of probability distributions. *Amsterdam Math. Institute.*