SIMULTANEOUS CONFIDENCE INTERVAL ESTIMATION[1]

by

S. N. Roy and R. C. Bose
Institute of Statistics, University of North Carolina

Institute of Statistics
Mimeograph Series No. 67
Limited Distribution
April, 1953

---

1. This research was supported in part by THE UNITED STATES AIR FORCE under Contract AF 18(600)-83 monitored by the Office of Scientific Research.

SIMULTANEOUS CONFIDENCE INTERVAL ESTIMATION

by S. N. Roy and R. C. Bose[1]

Institute of Statistics, University of North Carolina

Summary. The work of Neyman on confidence limits and of Fisher on fiducial limits is well known. However, in most applications the interval or limits for only a single parameter or a single function of the parameters has been considered. Recently Scheffe [2] and Tukey [3] have considered special cases of what may be called problems of simultaneous estimation, in which one is interested in giving confidence intervals for a set of finite or an infinite number of parametric functions such that the probability of parametric functions of the set being simultaneously covered by the corresponding intervals is a preassigned number $1 - \alpha$ $(0 < \alpha < 1)$.

In this paper we discuss in section 1, a set of sufficient conditions under which such simultaneous estimation is possible, and bring out the connection of this with a method of test construction considered by one of the authors in a previous paper [1].

In section 2 some univariate examples (including the ones due to Scheffe and Tukey) are considered from this point of view. The rest of the paper is concerned with multivariate applications, which are all believed to be new. The associated tests all turn out to be the same as

1. This work was supported in part by The United States Air Force under contract AF 18(600)-83 monitored by the Office of Scientific Research.

as in [1], except for the example in § 4.3, which is a multivariate generalization of Tukey's example (§ 2.2).

1. Introductory remarks on simultaneous estimation

§ 1.1. Let $y = (y_1, y_2, \ldots, y_n)$ be an observed set of random variables, whose joint distribution depends on unknown parameters, $\theta_1, \theta_2, \ldots, \theta_m$. Let

$$\pi_k = f_k(\theta_1, \theta_2, \ldots, \theta_m) \tag{1.1.1}$$

be a set of functions of the parameters, where the index k belongs to a finite or infinite set Γ . We shall consider the problem of making simultaneous confidence statements

$$\phi_{k1}(y_1, y_2, \ldots, y_n) \leq \pi_k \leq \phi_{k2}(y_1, y_2, \ldots, y_n) \tag{1.1.2}$$

with confidence coefficient $1 - \alpha$, which gives the probability that the statements (1.1.2) are simultaneously true for all $k \varepsilon \Gamma$.

This problem can be solved under the following circumstances. Suppose it is possible to find a set of functions

$$\psi_k(y_1, y_2, \ldots, y_n, \pi_k), \qquad k \varepsilon \Gamma \tag{1.1.3}$$

such that

$$d_1 \leq \psi_k \leq d_2, \qquad k \varepsilon \Gamma \tag{1.1.4}$$

implies (1.1.2) and conversely where d_1 and d_2 are constants independent of k. For a given $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ let

$$(1.1.5) \qquad W_\theta = \bigcap_k \left\{ d_1 \leq \psi_k \leq d_2 \right\}, \qquad k \varepsilon \underline{(\overline{\ })}$$

be the intersection of the regions (1.4.1). If W_θ is a Borel set for each admissible θ, and

$$(1.1.6) \qquad \text{Prob} \left\{ y \varepsilon W_\theta \,\middle|\, \theta \right\} = 1 - \alpha, \quad 0 < \alpha < 1$$

is independent of the parameters, then $1 - \alpha$ is also the chance that the statements (1.1.2) are simultaneously true for all $k \varepsilon \underline{(\overline{\ })}$.

<u>Proof</u>. If the sample point $y = (y_1, y_2, \ldots, y_n)$ belongs to W_θ, then (1.1.4) is true for every $k \varepsilon \underline{(\overline{\ })}$, and the same therefore holds for (1.1.2). Conversely if (1.1.2) is true for all $k \varepsilon \underline{(\overline{\ })}$, then the same holds for (1.1.4). Consequently the sample point y belongs to W_θ. Thus the statements (1.1.2) are simultaneously true when and only when $y \varepsilon W_\theta$, and the probability for this is by hypothesis $1 - \alpha$.

<u>Remark</u>. Let $F_\theta^{(1)}(y_1, y_2, \ldots, y_n)$ and $F_\theta^{(2)}(y_1, y_2, \ldots, y_n)$ be the supremum and infimum of ${}_k$ with respect to the variation of k over all elements of $\underline{(\overline{\ })}$. Then

$$(1.1.7) \qquad W_\theta = \left\{ d_1 \leq F_\theta^{(1)} \leq F_\theta^{(2)} \leq d_2 \right\}$$

provided that for each $y = (y_1, y_2, \ldots, y_n)$ there exist corresponding k_1 and k_2 such that $\psi_{k_1} = F_\theta^{(1)}$, $\psi_{k_2} = F_\theta^{(2)}$, (i. e. the suprema and infima are actually attained).

§ 1.2. Let H_0 be a hypothesis regarding the parameters, which fixes the value of $\pi_k = f_k(\theta_1, \theta_2, \ldots, \theta_m)$ for all $k\varepsilon(\underline{\ })$. Thus let $\pi_k = \pi_{k0}$ for $k\varepsilon(\underline{\ })$ if H_0 is true. Conversely let $\pi_k = \pi_{k0}$ for all $k\varepsilon(\underline{\ })$ imply the truth of H_0. Then a test of the hypothesis H_0 is obtained by rejecting H_0 when and only when, at least one of the statements

$$(1.2.1) \qquad \phi_{k1}(y_1, y_2, \ldots, y_n) \leq \pi_{k0} \leq \phi_{k2}(y_1, y_2, \ldots, y_n), \; k\varepsilon(\underline{\ })$$

if false. It is evident that the size of the test is α, since $1 - \alpha$ is the chance for the statements (1.2.1) to be simultaneously true. The region W_{θ_0} remains the same for all sets of parameters $\theta_0 = (\theta_{01}, \theta_{02}, \ldots, \theta_{0m})$ for which H_0 is satisfied. To calculate W_{θ_0} we can therefore take any set of values for the parameters consistent with H_0. The critical region for rejecting H_0 is then $\overline{W}_{\theta_0}$, the complement of W_{θ_0}. The power of the test against an alternative H for which the set of parameters is $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is

$$(1.2.2) \qquad 1 - \text{Prob}\left\{ y\varepsilon W_{\theta_0} \,\middle|\, \theta \right\} .$$

2. Applications to univariate simultaneous estimation problems.

§ 2.1. Let $y_1, y_2, \ldots, y_n$ be independent normal variates with common variance σ^2 (unknown), and let

$$E(y_i) = a_{i1}\theta_1 + a_{i2}\theta_2 + \ldots + a_{im}\theta_m, \qquad i = 1, 2, \ldots, n$$

where $\theta_1, \theta_2, \ldots, \theta_m$ are unknown parameters, and

$$n > \text{rank } (a_{ij}) = n_0 \text{ (say)}.$$

Clearly $n_0 \leq m$. A linear function π of the parameters $\theta_1, \theta_2, \ldots, \theta_m$ is said to be linearly estimable, if there exists a linear function Y of the variates such that

$$E(Y) = \pi \ .$$

In this case Y is said to be an unbiased linear estimate of π. The unbiased linear estimate with the minimum variance is called the best linear estimate of π.

Consider the problem of simultaneous estimation for a set of linear functions

$$\pi_k = \ell_{k1}\theta_1 + \ell_{k2}\theta_2 + \ldots + \ell_{km}\theta_m \tag{2.1.1}$$

such that the coefficient vectors $(\ell_{k1}, \ell_{k2}, \ldots, \ell_{km})$ form a vector space V_1 of rank $n_1 \leq n_0$, given an independent estimate s^2 of σ^2 based on n_2 degrees of freedom. Let

$$Y_k = c_{k1}y_1 + c_{k2}y_2 + \ldots + c_{kn}y_n \tag{2.1.2}$$

be the best linear estimate of π_k. Then the coefficient vectors $(c_{k1}, x_{k2}, \ldots, c_{kn})$ form a vector space V of rank n_1, and it is possible to choose n_1 mutually orthogonal vectors

$$(g_{i1}, g_{i2}, \ldots, g_{im}), \qquad i = 1, 2, \ldots, n_1$$

of unit length lying in V. If we set

$$U_i = g_{i1}y_1 + g_{i2}y_2 + \dots + g_{in}y_n, \qquad i = 1, 2, \dots, n_1$$

$$E(U_i) = \Phi_i. \qquad\qquad i = 1, 2, \dots, n_1$$

then there exist constants $b_{k1}, b_{k2}, \dots, b_{kn_1}$ such that

$$Y_k = b_{k1}U_1 + b_{k2}U_2 + \dots + b_{kn_1}U_{n_1}$$

$$\pi_k = b_{k1}\Phi_1 + b_{k2}\Phi_2 + \dots + b_{kn_1}\Phi_{n_1} .$$

Conversely each set of constants $b_{k1}, b_{k2}, \dots, b_{kn_1}$ determines a unique π_k and Y_k belonging to (2.1.1) and (2.1.2) respectively, so that the index k is in (1,1) correspondence with the set $(b_{k1}, b_{k2}, \dots, b_{kn_1})$.

Also $U_1, U_2, \dots, U_{n_1}$ are independently distributed normal variates with variance σ^2 and

$$\mathrm{Var}\ (Y_k) = (b_{k1}^2 + b_{k2}^2 + \dots + b_{kn_1}^2)\ \sigma^2$$

Since s^2 is an independent estimate of σ^2,

$$\mathrm{Est.\ var}\ (Y_k) = (b_{k1}^2 + b_{k2}^2 + \dots + b_{kn_1}^2)\ s^2$$

Let us set

(2.1.3)
$$k = \frac{Y_k - \pi_k}{\mathrm{Est.\ var}\ (Y_k)}$$

$$= \frac{b_{k1}(U_1 - \Phi_1) + b_{k2}(U_2 - \Phi_2) + \dots + b_{kn_1}(U_{n_1} - \Phi_{n_1})}{s\sqrt{b_{k1}^2 + b_{k2}^2 + \dots + b_{kn_1}^2}}$$

then

(2.1.4) $$-d \leq \psi_k \leq d$$

implies

$$Y_k - d\sqrt{\text{Est. var } (Y_k)} \leq \eta_k \leq Y_k + d\sqrt{\text{Est. var } (Y_k)}$$

since

$$\text{Sup}_k \ \psi_k = + \left\{ \sum_{i=1}^{n_1} (U_i - \bar{\emptyset}_i)^2/s^2 \right\}^{1/2}$$

and

$$\text{Inf}_k \ \psi_k = - \left\{ \sum_{i=1}^{n_1} (U_i - \bar{\emptyset}_i)^2/s^2 \right\}^{1/2}$$

it follows from the remark at the end of § 1.1., that W_θ the intersection of the regions (2.1.4) is given by

(2.1.5) $$W_\theta = \left\{ \sum_{i=1}^{n_1} (U_i - \bar{\emptyset}_i)^2/s^2 \leq d^2 \right\}$$

Now $\Sigma(U_i - \bar{\emptyset}_i)^2/n_1 s^2$ is distributed as F with degrees of freedom n_1, n_2. Hence if we put

$$d = \sqrt{n_1 F_\alpha(n_1, n_2)}$$

where $F_\alpha(n_1,n_2)$ is the upper α-point of the F-distribution with n_1, n_2 degrees of freedom, then the chance for $y_1, y_2, \ldots, y_n$ to lie in W_θ is $1 - \alpha$. Hence we get the simultaneous confidence intervals

(2.1.6) $$Y_k - \sqrt{n_1 F_\alpha(n_1, n_2) \times \text{Est. var } (Y_k)} \leq \eta_k$$

$$\leq Y_k + \sqrt{n_1 F_\alpha(n_1, n_2) \times \text{Est. var } (Y_k)}$$

with confidence coefficient $1 - \alpha$, for the set of parametric functions (2.1.1). This is essentially Scheffe's [2] result when expressed in the general linear form. It should be noted that the confidence intervals (2.1.6) are independent of the linear functions U_i.

Again suppose we wish to test the hypothesis H_0, that any n_1 independent linear functions belonging to the set (2.1.1) vanish. This is equivalent to the vanishing of $\bar{\emptyset}_i$, $i = 1, 2, \ldots, n_1$. It follows from § 1.2., that a test of the hypothesis H_0 is obtained by using the region of rejection

(2.1.7) $$\sum_{i=1}^{n_1} U_i^2/n_1 s^2 > F_\alpha(n_1,n_2)$$

Thus we get the usual F-test of the hypothesis H_0.

§ 2.2. Let $y_1, y_2, \ldots, y_n$ be normal variates for which

(2.2.1) $$E(y_i) = \theta_i,\ var(y_i) = \sigma^2 \quad i = 1, 2, \ldots n$$

(2.2.2) $$cov(y_i,y_j) = \rho\sigma^2 \quad i,j=1,2,\ldots,n,\ i \neq j$$

where ρ is known, m_i and σ^2 are unknown, but an independent estimate s^2 of σ^2 based on n' degrees of freedom is available. It is required to obtain a simultaneous estimate of the mean differences

(2.2.3) $$\theta_i - \theta_j \qquad i, j=1,2, \ldots, n, \quad i \neq j$$

In contradistinction to the example considered in § 2.1., we have now a finite set of parametric functions.

Let

$$z_i + \ell\bar{\theta} = y_i + \ell\bar{y}$$

where

$$\bar{y} = (y_1+y_2+\ldots+y_n)/n, \; \bar{\theta} = (\theta_1+\theta_2+\ldots+\theta_n)/n$$

and the disposable constant ℓ is so adjusted that the z_i's are uncorrelated. Then

$$\text{(2.2.4)} \qquad E(z_i) = \theta_i, \; \text{var}\,(z_i) = \sigma^2(1 - \rho) \qquad i = 1,2,\ldots,n$$

Let

$$\text{(2.2.5)} \qquad \psi_{ij} = \frac{(z_i-\theta_i)-(z_j-\theta_j)}{s\sqrt{1-\rho}} \qquad i, j = 1, 2, \ldots, n, \; i \neq j \; .$$

Then

$$\text{(2.2.6)} \qquad |\psi_{ij}| \leq d$$

implies

$$\text{(2.2.7)} \qquad y_i-y_j-ds\sqrt{1-\rho} \leq \theta_i-\theta_j \leq y_i-y_j+ds\sqrt{1-\rho}$$

Let W_θ be the intersection of the regions (2.2.6). Then clearly the necessary and sufficient condition for the sample point to lie in W_θ is that

$$\text{(2.2.8)} \qquad q = \frac{w}{s\sqrt{1-\rho}} \leq d$$

where

(2.2.9) $$w = \mathrm{Sup}_i \left| z_i - \theta_i \right| \qquad i = 1, 2, \ldots, n$$

Thus if we set $d = q_\alpha(n,n')$, where $q_\alpha(n,n')$ is the upper α-point of the distribution of the ratio of the range of n independent normal variates with zero mean, to the square root of an independent estimate of their common variance based on n' degrees of freedom, then the required simultaneous confidence intervals for the parametric functions (2.2.3) are

(2.2.10) $$y_i - y_j - sq_\alpha(n,n')\sqrt{1-\rho} \leq \theta_i - \theta_j \leq y_i - y_j + sq_\alpha(n,n')\sqrt{1-\rho}$$

This result is due to Tukey [3]. In particular $y_1, y_2, \ldots, y_n$ may be means of n random samples of equal size drawn from normal populations with a common (unknown) variance, or may be the estimated treatment effects in a randomized block, or a balanced incomplete block experiment.

We can test the hypothesis H_0 that

(2.2.11) $$\theta_1 = \theta_2 = \ldots = \theta_n$$

by using as the region of rejection

(2.2.12) $$\frac{R}{s\sqrt{1-\rho}} > q_\alpha(n,n')$$

where

$$R = \mathrm{Sup}_{i,j} \left| y_i - y_j \right|$$

is the range of the random variates $y_1, y_2, \ldots, y_n$. Thus we arrive at a test different from the classical analysis of variance test.

§ 2.3. In factorial experiments we are usually interested in estimating linear functions of treatment effects, whose estimates are independently and normally distributed with a common variance, which can be independently estimated by an appropriate multiple of the error mean square in the analysis of variance. The distribution needed for simultaneous estimation in this case, is slightly different from that occuring in § 2.2.

Suppose, for example, that we have observations for a 2x2x2x2 factorial experiment with factors A, B, C, D, and that we are interested in simultaneously estimating the main effects and two factor interactions only. We shall suppose that the experiment is so laid out that none of these is confounded in any replication. Let t_{11}, t_{22}, t_{33}, t_{44} denote the true main effects and t_{12}, t_{13}, t_{14}, t_{23}, t_{24}, t_{34} the true two factor interactions. The order of the subscripts in t_{ij} is immaterial, i.e., $t_{ij} = t_{ji}$. We can then write in the usual notation

(2.3.1) $$t_{11} = \frac{1}{8}\,(a-1)(b+1)(c+1)(d+1)$$

(2.3.2) $$t_{12} = \frac{1}{8}\,(a-1)(b-1)(c+1)(d+1)$$

with similar expressions for other main effects and interactions. Let y_{ij} be the estimate of t_{ij}. Then reasoning as before we get the following simultaneous confidence intervals for t_{ij}:

(2.3.) $$y_{ij} - x_\alpha(n,n')\sqrt{\text{Est.var}(y_{ij})} \leq t_{ij}$$

$$\leq y_{ij} + x_\alpha(n,n') \sqrt{\text{Est.var}(y_{ij})}$$

where there are n' degrees of freedom available for the estimate of error, and $n = 10$ is the number of linear functions to be estimated.

The meaning of $x_\alpha(n,n')$ is as follows: Let $x_1, x_2, \ldots, x_n$ be independent normal variates with zero mean and variance σ^2. Let $|x|$ be the maximum of $|x_1|$, $|x_2|$, ..., $|x_n|$ and let s^2 be an independent estimate of σ^2 based on n' degrees of freedom. Then $x_\alpha(n,n')$ is the upper α-point of the distribution of $|x|/s$.

A test of the hypothesis H_0 that all the linear functions t_{ij} to be estimated are simultaneously zero, is obtained by using as the region of rejection

$$\text{Sup}_{i,j} |y_{ij}| \geq x_\alpha(n,n') \sqrt{\text{Est.var}(y_{ij})} \tag{2.3.4}$$

In a factorial experiment in which each factor is at more than two levels, the above result will still apply if the n linear functions to be simultaneously estimated (or tested for vanishing) are so chosen that their estimates are independently distributed with a common variance.

3. Notation and preliminaries for multivariate applications.

As far as possible Greek letters will stand for population parameters and Roman letters over the first half of the alphabet for given (non-stochastic) quantities and over the latter part from, say, s to the and for sample quantities, capital letters for matrices, small letters for scalars, small letters underscored for column vectors and such letters underscored and primed will stand for row vectors. Some exceptions to this, which are unavoidable, will be clearly indicated at the proper places. As usual the transpose of a matrix or a column vector will be denoted by priming such quantities. The absolute value of the determinant of a square matrix M will be denoted by $|M|$ and the absolute value of a scalar m by $|m|$. To indicate the structure, a p x q matrix, say M, or a p x 1 column vector, say $\underline{m}$, will sometimes be written respectively as M(p x q) or $\underline{m}$(p x 1). A positive definite matrix will be called a p.d. matrix and a positive semi-definite matrix a p.s.d. matrix. 'Almost everywhere', i.e., 'except for a set of (probability) measure zero' will be referred to as a.e. A matrix B whose typical element is b_{ij} will sometimes be denoted by (b_{ij}). A diagonal matrix whose diagonal elements are, say, $a_1, a_2, \ldots, a_p$ will be denoted by D_a. $\tilde{M}$ or $\tilde{M}$(p x p) will stand for a triangular matrix with a configuration given by:

$$\tilde{M} = \begin{pmatrix} m_{11} & 0 & 0 \ldots\ldots 0 \\ m_{21} & m_{22} & 0 \ldots\ldots 0 \\ \cdot & \cdot & \cdot \ldots\ldots 0 \\ m_{p2} & \cdot & \cdot \ldots\ldots m_{pp} \end{pmatrix}$$

A normal variate x with mean ξ and variance σ^2 will be called $N(\xi, \sigma^2)$. A column vector $\underline{x}$(p x 1) whose components have a p-variate normal distribution about a mean vector $\underline{\xi}$(p x 1) and with a covariance matrix $\Sigma(p \times p)$ will be called $N(\underline{\xi}, \Sigma)$. This Σ is a symmetric and always at least a p.s.d. matrix. In the problems we shall be discussing in this paper this Σ will be assumed to be p.d. A random sample of $(n+1)$ individuals from a $N(\underline{\xi}, \Sigma)$, i.e., a matrix $X(p \times (n+1))$ will have the probability density

$$\left[1/(2\pi)^{p(n+1)/2}\right] \Sigma^{(n+1)/2} \text{ Exp}\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(X-\xi)(X'-\xi')\right]$$

where ξ ($p \times (n+1)$) stands for $(\overset{n+1}{\underline{\xi}\underline{\xi}\ldots\ldots\underline{\xi}})p$. Notice that in the matrix X any element in the i-th row and j-th column is to be called x_{ij} where $i = 1,2,\ldots,p$ and $j = 1,2,\ldots, n+1$ and where i stands for a variate and j for an individual. A matrix X having the above probability law will be called an X: $N(\underline{\xi}, \Sigma)$. Also let $\bar{x}_i \equiv \sum_{j=1}^{n+1} x_{ij}/(n+1)$ $(i = 1,2,\ldots,p)$ and let $\underline{x}' \equiv (\bar{x}_1 \ldots \bar{x}_p)$. It is well known that by an orthogonal transformation we can change over from $X(p \times (n+1))$ to $(\underset{n}{Y} \;\; \sqrt{n+1}\,\underset{1}{\underline{x}})p$, where $YY' = nS(p \times p) = XX' - (n+1)\underline{x}\underline{x}'$, S being the sample covariance matrix, and where Y and $\underline{x}$ have the joint probability density

$$\text{Const. Exp}\left[-\tfrac{1}{2}\operatorname{tr}\; \Sigma^{-1} \;\; YY' + (n+1)(\underline{x}-\underline{\xi})(\underline{x}'-\underline{\xi}')\right]$$

while Y and $\underline{x}$ have the respective probability densities

$$\text{Const. Exp}\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}YY'\right] \text{ and}$$

$$\text{Const. Exp}\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\;(n+1)(\underline{x}-\underline{\xi})(\underline{x}'-\underline{\xi}')\right]$$

An $Y(p \times n)$ having the above probability can obviously be called $Y{:}N(\underline{0},\Sigma)$. For problems on covariance matrices or canonical oorrelations or regressions we shall start not from $X(p \times (n+1))$: $N(\underline{\xi}, \Sigma)$, but directly from

$Y(p \times n)$: $N(\underline{0},\Sigma)$. As is well known there is a lot of arbitrariness in Y, but this does not matter at all in the results we are ordinarily interested in, because all such results ultimately come out in terms of $\underline{x}$ and YY', i.e., S. In sections 3, 4 and 5 of this paper which, in a certain sense, are a follow-up of a previous paper [1] repeated use is made of the fact that if $\underline{x}(p \times 1)$ is $N(\underline{\xi}, \Sigma)$, then, for a fixed, i.e., non-stochastic $\underline{a}(p \times 1)$, the scalar $\underline{a}'\underline{x}$ is $N(\underline{a}'\underline{\xi}, \underline{a}'\Sigma\underline{a})$, and thus multivariate problems are thrown back on univariate and bivariate problems exactly in the same manner as in the previous paper. We also make repeated use of the result that $\mathrm{tr}\, A(p \times q)\, B(q \times p) = \mathrm{tr}\, BA$.

4. Multivariate estimation and testing problems on means.

In three subsections under this section we shall consider three estimation problems each coupled with a corresponding problem in testing of hypothesis. It will be evident from the titles to the subsections that the first problem is, in a sense, a special case of the second and the second again of the third. But for expository purposes and from considerations of practical usefulness there is an obvious advantage in discussing the three cases separately and in order of increasing generality and difficulty. It may be also noted that so far as testing of hypotheses is concerned, out of the three major problems considered in § 4.1, § 4.2 and § 4.5 of this section the last two have been already discussed in a previous paper [1] and the actual tests offered there are precisely the same as are offered here by inverting the confidence estimation procedures.

§ 4.1. Estimation and testing problem on $\underline{\xi}$ from an $N(\underline{\xi}, \Sigma)$.

Given an $X(p \times (n+1)): N(\underline{\xi}, \Sigma)$, suppose we try to obtain simultaneous confidence bounds on arbitrary linear compounds of the population mean vector $\underline{\xi}$. Now consider the statement that

$$(n+1)^{1/2} \left| \underline{a}'(\underline{x} - \underline{\xi}) \right| / (\underline{a}'S\underline{a})^{1/2} \leq c ,$$

or

$$(4.1.1) \qquad (n+1) \quad \underline{a}'(\underline{x} - \underline{\xi})(\underline{x}' - \underline{\xi}')\underline{a}/\underline{a}'S\underline{a} \leq c^2 ,$$

where $\underline{x}$ is the sample mean vector and S is the sample covariance matrix, already defined in section 3. , and $\underline{a}(p \times 1)$ is an arbitrary non-null non-stochastic column vector and c is a given positive constant. The statement (4.1.1) stems from the customary Student's t-test and the associated confidence interval (both having well-known optimum properties) relating to the parameter. $\underline{a}'\underline{\xi}$. Now, for a given (positive) c and given $\underline{x}$, $\underline{\xi}$, S and of course n, the set of all statements (3.1.1) for all possible non-null vectors $\underline{a}$ is exactly equivalent to the statement that

$$(4.1.2) \qquad \operatorname{Sup}_{\underline{a}} (n+1)\underline{a}'(\underline{x} - \underline{\xi})(\underline{x}' - \underline{\xi}') \, \underline{a}/\underline{a}'S\underline{a} \leq c^2 ,$$

the "Sup" being with reference to variation over $\underline{a}$. It is well known that this "Sup" comes out as $\operatorname{tr}(n+1)S^{-1}(\underline{x} - \underline{\xi})(\underline{x}' - \underline{\xi}')$, i. e.,

$$\operatorname{tr}(n+1)(\underline{x}' - \underline{\xi}')S^{-1}(\underline{x} - \underline{\xi})$$

or simply as

$$(n+1)(\underline{x}' - \underline{\xi}')S^{-1}(\underline{x} - \underline{\xi}) .$$

It is also well known that when the true population mean vector $\underline{\xi}^* = \underline{\xi}$, then this is distributed as the central Hotelling's T^2 with D. F. (p) and (n+1-p) and when $\underline{\xi}^* \neq \underline{\xi}$ this is distributed as the non-central Hotelling's T^2 with the same D. F. and with the non-centrality parameter

$$\Upsilon^2 = (\underline{\xi}^{*\prime} - \underline{\xi}') \, \Sigma^{-1}(\underline{\xi}^* - \underline{\xi}).$$

Going back to (4.1.1) it is thus easy to see that if, for all $\underline{\xi}$ and all non-null $\underline{a}$,

$$(4.1.3) \qquad P \left[\frac{(n+1)\underline{a}'(\underline{x} - \underline{\xi})(\underline{x}' - \underline{\xi}')\underline{a}}{\underline{a}'S\underline{a}} \leq c^2 \quad \underline{\xi}^* = \underline{\xi} \right] = 1 - \alpha ,$$

then $c^2 = T^2_\alpha$ is the upper α-point of the central Hotelling's T^2-distribution with D. F. p and (n+1-p) and can be conveniently written as $T^2_\alpha(p, n+1-p)$. From (4.13) we have thus, with a confidence coefficient $1 - \alpha$, the set of simultaneous or multiple confidence bounds (for all $\underline{\xi}$ and all non-null $\underline{a}$):

$$(4.1.4) \qquad \underline{a}'\underline{x} - \left[T^2_\alpha(p,\ n+1-p)(\underline{a}'S\underline{a})/n+1 \right]^{1/2} \leq \underline{a}'\underline{\xi}$$

$$\leq \underline{a}'\underline{x} + \left[T^2_\alpha(p,\ n+1-p)(\underline{a}'S\underline{a})/n+1 \right]^{1/2} .$$

It should be noted that (4.1.4) gives the simultaneous confidence bounds on all arbitrary linear compounds of the p components of the population mean vector $\underline{\xi}$. The shortness (in the sense of probability) of this set of confidence bounds, i. e., the probability of these bounds covering $\underline{\xi}$ when, in fact, $\underline{\xi}^* \neq \underline{\xi}$, is obviously

$$1 - P \left[\text{non-central } T^2 \geq T^2_\alpha(p,\ n+1-p) \,\middle|\, \Upsilon^2 \right].$$

From the well-known fact that the power function of Hotelling's T-test is a monotonically increasing function of the non-negative τ , it follows, therefore, that the shortness of the confidence bound (4.1.4) tends to zero as $\tau \longrightarrow \infty$.

From § 1.2. the critical region of the associated hypothesis: $\underline{\xi} = \underline{\xi}_0$, i.e., of the hypothesis: $\bigcap_{\underline{a}}(\underline{a}'\underline{\xi} = \underline{a}'\underline{\xi}_0)$ turns out to be:

$$(n+1)(\underline{x}' - \underline{\xi}'_0)S^{-1}(\underline{x} - \underline{\xi}_0) \geq T^2_\alpha ,$$

which implies that, for at least one $\underline{a}$, the set of confidence bounds (2.1.4) does not include $\underline{a}'\underline{\xi}_0$; the region of acceptance based on the opposite inequality will imply that, for all $\underline{a}$, the set of bounds (4.1.4) will include $\underline{a}'\underline{\xi}_0$. Mention has already been made of the monotonic character of the power function of the test, with respect to the deviation or non-centrality parameters.

§ 4.2. Estimation and testing problem on mean differences from

$$N(\underline{\xi}_i, \Sigma)(i = 1, 2, \ldots, k) .$$

Given $X_i(p \times (n_i+1)): N(\underline{\xi}_i, \Sigma)$ $(i = 1, 2, \ldots, k)$ let us try to obtain a set of simultaneous confidence bounds on all arbitrary double linear compounds of the p-components of the k population mean vectors measured from the weighted grand mean vector. Consider now the statement

$$(4.2.1) \qquad \sum_{i=1}^{k} b_i\underline{a}'(n_i+1)^{1/2}(\underline{x}_i-\underline{x}-\underline{\xi}_i+\underline{\xi}) \quad \leq \left[(k-1)c^2\underline{a}'S\underline{a}\right]^{1/2}$$

where $\underline{x}_i$ is the mean vector for the i-th sample,

$$\underline{x} = \sum_{i=1}^{k}(n_i+1)\underline{x}_i / \sum_{i=1}^{k}(n_i+1), \quad \underline{\xi} = \sum_{i=1}^{k}(n_i+1)\underline{\xi}_i / \sum_{i=1}^{k}(n_i+1),$$

and whereS is the pooled "within" covariance matrix of the k-samples, given by

$$(\sum_{i=1}^{k} n_i)S = \sum_{i=1}^{k} [X_i X_i' - (n_i+1)\underline{x}_i\underline{x}_i'],$$

and c is a given (positive) constant, $\underline{a}(p \times 1)$ is an arbitrary non-null non-stochastic column vector and b_i's are arbitrary coefficients subject to $\sum_{i=1}^{k} b_i^2 = 1$.

If we now use the result that

$$\text{with } \sum_{i=1}^{k} b_i^2 = 1,$$

$$\sum_{i=1}^{k} b_i y_i \leq \pm\sqrt{d^2} \quad \begin{array}{c}\longrightarrow\\ \longleftarrow\end{array} \quad \sum_{i=1} y_i^2 \leq d^2 \quad ,$$

(then) it directly follows that, given all the other quantities, and under all possible variations of b_i's subject to $\sum_{i=1}^{k} b_i^2 = 1$, the statement (4.2.1) is precisely equivalent to the statement that

$$\sum_{i=1}^{k} [\underline{a}'(n_i+1)^{1/2}(\underline{x}_i-\underline{x}-\underline{\xi}_i+\underline{\xi})]^2/(k-1)\underline{a}'S\underline{a} \leq c^2,$$

or

$$(4.2.2) \quad \sum_{i=1}^{k} \underline{a}'(n_i+1)(\underline{x}_i-\underline{x}-\underline{\xi}_i+\underline{\xi})(\underline{x}_i'-\underline{x}'-\underline{\xi}_i'+\underline{\xi}')\underline{a}/(k-1)\underline{a}'S\underline{a} \leq c^2.$$

Now putting

$$(k-1)S^* = \sum_{i=1}^{k} (n_i+1)(\underline{x}_i-\underline{x}-\underline{\xi}_i+\underline{\xi})(\underline{x}_i'-\underline{x}'-\underline{\xi}_i'+\underline{\xi}'),$$

the statement (4.2.2), for all possible values of the non-null $\underline{a}$, is precisely equivalent to:

$$(4.2.3) \qquad \mathrm{Sup}_{\underline{a}} [\underline{a}'S^*\underline{a}/\underline{a}'S\underline{a}] \leq c^2 \quad ,$$

As observed in a previous paper [1] S is, a. e., p.d. and S^* is a. e., p.s.d. of rank $q = \min(p, k-1)$(p.s.d. if $p > k-1$ and p.d. if $p \leq k-1$) and that the $\mathrm{Sup}_{\underline{a}} [\underline{a}'S^*\underline{a}/\underline{a}'S\underline{a}]$ is just the largest root θ_q of the p-th degree determinantal equation in θ: $|S^* - \theta S| = 0$, of which all roots are non-negative, p-q of them always zero and q are, a.e., positive. Thus (4.2.3), i.e., (4.2.2), i. e., (4.2.1) under all permissible variations of $\underline{a}$ and b_i's, turns out to be equivalent to:

$$(4.2.4) \qquad \theta_q \leq c^2 \quad .$$

The distribution of this θ_q on the null hypothesis, i.e., when the true population means $\underline{\xi}_i^*$'s = $\underline{\xi}_i$'s . is known and relatively easy and involve as parameters p, k-1, $\sum_{i=1}^{k} n_i$. Computation of the 5 % and 1 % points is also under way. Thus if

$$(4.2.5) \qquad P [\theta_q \leq \theta_\alpha \mid \text{true population means} = \xi_i\text{'s}] = 1 - \alpha \quad ,$$

we can write $\theta_\alpha = \theta_\alpha(p, k-1, \sum_{i=1}^{k} n_i)$, and now combining (4.2.1)-(4.2.5), we have, with a confidence coefficient $1 - \alpha$, the following set of multiple confidence statements (for all $\underline{\xi}_i$'s, all non-null $\underline{a}$'s and all b_i's subject to $\sum_{i=1}^{k} b_i^2 = 1$, and to accomplish equality, also $\sum_{i=1}^{k} b_i = 1$):

$$(4.2.6) \qquad \sum_{i=1}^{k} b_i \underline{a}'(n_i+1)^{1/2}(\underline{x}_i-\underline{x}) - [(k-1)\theta_\alpha \underline{a}'S\underline{a}]^{1/2}$$

$$\leq \sum_{i=1}^{k} b_i \underline{a}'(n_i+1)^{1/2}(\underline{\xi}_i - \underline{\xi}) \leq \sum_{i=1}^{k} b_i \underline{a}'(n_i+1)^{1/2}(\underline{x}_i-\underline{x})$$

$$+ [(k-1)\theta_\alpha \underline{a}'S\underline{a}]^{1/2} \quad ,$$

where $\theta_\alpha = \theta_\alpha(p, k-1, \sum_{i=1}^{k} n_i)$. It may be observed that (4.2.6) gives the simultaneous confidence bounds on all arbitrary double linear compounds of the p components of the difference between the k population mean vectors $\underline{\xi}_i$'s and the weighted grand mean of these which is $\underline{\xi}$.

To discuss the shortness of (4.2.6) consider the non-central distribution of θ_q of (4.2.6), i.e., when not all the true population means $\underline{\xi}_i^*$'s = $\underline{\xi}_i$'s. This distribution is extremely difficult but is well known to involve as parameters, besides the D. F.'s, the positive roots Θ_1, $\Theta_2, \ldots, \Theta_s$ ($s \leq \min(p, k-1)$) of the determinantal equation in Θ : $\left|\Sigma^* - \Theta\Sigma\right| = 0$. Here Σ is the common covariance matrix of the k populations and $\Sigma^* = \sum_{i=1}^{k} (n_i+1)(\underline{\xi}_i^*-\underline{\xi}^*-\underline{\xi}_i+\underline{\xi})(\underline{\xi}_i^{*\prime}-\underline{\xi}^{*\prime}-\underline{\xi}_i'+\underline{\xi}')/(k-1)$. This Σ^* is necessarily at least p.s.d. of rank $\leq \min(p, k-1)$, = s(say), so that, out of the p roots of the equation in Θ, p-s are zero and s positive. If now we formally write

$$(4.2.7)\quad P[\theta_q \leq \theta_\alpha(p,\ k-1,\ \sum_{i=1}^{k} n_i) \mid \text{not all true population means } \underline{\xi}_i^{*}\text{'s}=\underline{\xi}_i\text{'s}]$$

$$=\psi(\alpha,\ p,\ k-1,\ \sum_{i=1}^{k} n_i,\ \Theta_1,\ \Theta_2,\ \ldots,\ \Theta_s),$$

then we note that while ψ is extremely difficult to obtain, there is a pretty good upper bound to it [1] given by

$$(4.2.8)\quad \psi < [P(\text{central } F \leq \theta_\alpha)]^{p-s} \prod_{i=1}^{s} P[\text{non-central } F \leq \theta_\alpha \mid \text{deviation parameter} = \Theta_i],$$

where all F's are on D. F. (k-1) and $\sum_{i=1}^{k} n_i$. Furthermore, as stated and proved elsewhere [1], this ψ is also a monotonically decreasing function of the deviation parameters and tends to zero as these tend to infinity.

With two populations (and samples), we have q = min(p, 1) = 1, and thus only one positive sample root to be called, say θ, and at the most one positive population root to be called, say Θ. It is easy to check that in this case

$$(4.2.9)\quad \theta = \frac{(n_1+1)(n_2+1)}{n_1+n_2+2}\, \mathrm{tr} S^{-1}(\underline{x}_1-\underline{x}_2-\underline{\xi}_1+\underline{\xi}_2)(\underline{x}_1'-\underline{x}_2'-\underline{\xi}_1'+\underline{\xi}_2'),$$

and

$$\Theta = \frac{(n_1+1)(n_2+1)}{n_1+n_2+2}\, \mathrm{tr}\Sigma^{-1}(\underline{\xi}_1^{*}-\underline{\xi}_2^{*}-\underline{\xi}_1+\underline{\xi}_2)(\underline{\xi}_1^{*\prime}-\underline{\xi}_2^{*\prime}-\underline{\xi}_1'+\underline{\xi}_2')$$

and it is well known that, on the null hypothesis, θ is distributed as central Hotelling T^2 with D. F. p and n_1+n_2+1-p, and on the non-null hypothesis as non-central Hotelling T^2 with the same D. F. and with a deviation parameter Θ. It is also easy to check that in this case the confidence statement (4.2.6) reduces to

$$(4.2.10) \qquad \underline{a}'(\underline{x}_1-\underline{x}_2) - \left[\frac{n_1+n_2+2}{(n_1+1)(n_2+1)} T_\alpha^2 \underline{a}'S\underline{a}\right]^{1/2} \leq \underline{a}'(\underline{\xi}_1-\underline{\xi}_2)$$

$$\leq \underline{a}'(\underline{x}_1-\underline{x}_2) + \left[\frac{n_1+n_2+2}{(n_1+1)(n_2+1)} T_\alpha^2 \underline{a}'S\underline{a}\right]^{1/2},$$

where $T_\alpha^2 = T_\alpha^2(p, n_1+n_2+1-p)$ is the upper α-point of Hotelling's T^2. The shortness of (4.2.10) which is now a degenerate form of (4.2.7) is exactly known and of course tends to zero as $\Theta \longrightarrow \infty$.

From § 1.2., the critical region of the associated hypothesis: $\underline{\xi}_1 = \underline{\xi}_2 = \dots = \underline{\xi}_k$, i. e., of the hypothesis $\bigcap_{\underline{a}}(\underline{a}'\underline{\xi}_i=\underline{a}'\underline{\xi})(i=1,2,\dots,K)$, turns out to be the same as given in a previous paper, namely:

$$(4.2.11) \qquad \phi_q \geq \theta_\alpha(p, k-1, \sum_{i=1}^{k} n_i)$$

with a power function $1 - \psi(\alpha, p, k-1, \sum_{i=1}^{k} n_i, \Phi_1, \dots, \Phi_s)$ where ϕ_q is the largest characteristic root of $S^{-1} \cdot \sum_{i=1}^{k} (n_i+1)(\underline{x}_i-\underline{x})(\underline{x}'-\underline{x}')/k-1$ and where the Φ's are the roots of the equation in Φ:

$$\left| \sum_{i=1}^{k} (n_i+1)(\underline{\xi}_i-\underline{\xi})(\underline{\xi}_i'-\underline{\xi}')/k-1 - \underline{\emptyset}\ \Sigma \right| = 0.$$

The properties of this power function, such as indicated under (4.2.8), have been already discussed in the previous paper [1].

§ 4.3. An important subset of the set of bounds (4.2.6).

Suppose now that, instead of all contrasts of the type:

$$\sum_{i=1}^{k} b_i \underline{a}'(n_i+1)^{1/2}(\underline{\xi}_i - \underline{\xi})$$

(with the given restrictions on $\underline{a}$ and b_i's), we are interested in contrasts of the type: $\underline{a}'(\xi_i - \underline{\xi}_j)$, for all non-null $\underline{a}'$ and all $i \neq j = 1, 2, \ldots, k$. It is easy to offer a multiple set of confidence bounds for contrasts of this type, which can be regarded as one kind of multivariate (under unequal sample sizes) analogue of a somewhat similar set given by Tukey for the corresponding univariate situations, and discussed in section 2. of this paper. The proposed set is built up as follows. With the same notation as before, and with $n_{ij} = (n_i+1)(n_j+1)/(n_i+n_j+2)$ note that

$$T_{ij}^2 = n_{ij} \cdot (\underline{x}_i' - \underline{x}_j' - \underline{\xi}_i' + \underline{\xi}_j)S^{-1} (\underline{x}_i - \underline{x}_j - \underline{\xi}_i + \underline{\xi}_j) =$$

largest value (under variation of $\underline{a}$) of:

$$n_{ij}\underline{a}' (\underline{x}_i - \underline{x}_j - \underline{\xi}_i + \underline{\xi}_j) (\underline{x}_i' - \underline{x}_j' - \underline{\xi}_i' + \underline{\xi}_j') \underline{a} / \underline{a}'S\underline{a} .$$

Thus, for a given pair (i, j), the statement that $T_{ij}^2 \leq T_\alpha^2$ is exactly equivalent to the statement that, for all non-null $\underline{a}$'s ,

$$\underline{a}'(\underline{x}_i-\underline{x}_j)-[T^2_\alpha\,\underline{a}'S\underline{a}/n_{ij}]^{1/2} \leq \underline{a}'(\underline{\xi}_i-\underline{\xi}_j) \leq \underline{a}'(\underline{x}_i-\underline{x}_j)+[T^2_\alpha\underline{a}'S\underline{a}/n_{ij}]^{1/2} .$$

We observe that when the true population means are $\underline{\xi}_i$'s, this T^2_{ij} is distributed as Hotelling's T^2 with D.F. p and $\sum_{i=1}^{k} n_i + 1 - p$.

Now, considering all pairs (i,j) out of k samples(and k populations), it is easy to see that the statement; all T^2_{ij}'s $\leq T^2_\alpha$, is precisely equivalent to the statement that the largest T^2_{ij} out of all pairs $\leq T^2_\alpha$, which again is equivalent to the statement that, for all non-null $\underline{a}$'s and all pairs (i, j) out of k,

$$(4.3.1)\quad \underline{a}'(\underline{x}_i-\underline{x}_j)-[T^2_\alpha\underline{a}'S\underline{a}/n_{ij}]^{1/2} \leq \underline{a}'(\underline{\xi}_i-\underline{\xi}_j) \leq \underline{a}'(\underline{x}_i-\underline{x}_j)+[T^2_\alpha\underline{a}'S\underline{a}/n_{ij}]^{1/2}.$$

If the confidence coefficient of (4.3.1) is to be $1-\alpha$, then $T_\alpha = T_\alpha(p, n_1\ n_2, \ldots, n_k)$ will be given by

$$(4.3.2)\qquad P\,[\text{ Largest } T^2_{ij} \text{ out of } (k_2) \text{ pairs} \geq T^2_\alpha \mid \text{true population means} = \underline{\xi}_i\text{'s}] = \alpha .$$

It will be obvious that the distribution of the largest T_{ij} involves as parameters just p and n_1, n_2, ..., n_k. It is easy to see that the distribution can be usable or manageable only when the number of parameters is small, for example when $n_1 = n_2 = \ldots = n_k$ and this case, with $p = 1$, in addition, becomes identical with the one considered in § 2.2. It may also be noted that when $k = 2$, this largest T^2_{ij} will of course be Hotelling's T^2 distributed with D. F. p and n_1+n_2+1-p. Also the shortness of the confidence bounds (4.3.1) can be formally written as

P [Largest T^2_{ij} out of $\binom{k}{2}$ pairs $\leq T^2_{\alpha}(p, n_1, n_2, \ldots, n_k)$ | not all true population means $\underline{\xi}^*_i$'s = $\underline{\xi}_i$'s].

It is important to observe that while each T_{ij} is individually distributed as a central Hotelling's T with D. F. p and $\sum_{i=1}^{k} n_i+1-p$ the $\binom{k}{2}$ T_{ij}'s are not independent, nor do we know what the distribution of the largest central T_{ij} is, to say nothing of the non-central case, so that we have not yet been able to reduce the confidence statement (4.3.1) to concrete terms the same way as we have found possible for the other cases discussed in this paper. The distribution problem arising in this situation is now under investigation.

For the associated problem of testing $H_0: \underline{\xi}_1 = \ldots = \underline{\xi}_k$, we set up as before, the rule that if, for all non-null $\underline{a}$ and all pairs (i, j), the bounds(4.3.1) include zero, we accept H_0 and reject it otherwise. The properties (including power) of this test is tied up in an obvious manner with those of the multiple confidence interval statement (2.3.1).

Notice that so far, in testing of hypothesis by inversion of confidence statements, we have considered two-decision problems. Suppose, at this point, for purposes of illustration, we offer a multi-decision procedure, namely that, for a given pair (i, j), we accept or reject $H(\underline{\xi}_i = \underline{\xi}_j)$ according as the set of bounds (4.3.1) for all $\underline{a}$ but for that pair (i, j) includes or excludes zero. It is obvious that in all the other situations considered so far we could set up similar multi-decision procedures. But no really adequate and complete (but only a very partial) treatment of such procedures can be possibly given within the set-up of this paper. A proper treatment can only be given in terms of either the Wald theory or some suitable modification of it.

4.4. Further observations. In many situations it might be of greater physical interest to be able to make, instead of (4.2.6) or even of (4.3.1), a set of just $p \times \binom{k}{2}$ confidence interval statements, each relating to just one variate and difference between one of $\binom{k}{2}$ pairs. In other words, if $\underline{\xi}_i' = (\xi_{1i}, \xi_{2i}, \ldots, \xi_{pi})$ $(i = 1, 2, \ldots, k)$ denote the p means for the i the population, then we would like to be able to make a statement like the following:

$$(4.4.1) \qquad f_{jii'}(X_1, X_2, \ldots, X_k) \le \xi_{ji} - \xi_{ji'} \le F_{jii'}(X_1, X_2, \ldots, X_k)$$

(with obvious applications to subsections 2.1 and 2.2), for all $i \neq i' = 1, 2, \ldots, k$ and all $j = 1, 2, \ldots, p$, where $f_{jii'}$ and $F_{jii'}$ are supposed to be two different functions of the whole set of $p \times \sum_{i=1}^{k} (n_i+1)$ raw observations. It is clear that (4.4.1) is a subset of (4.3.1) which again is a subset of (4.2.6). Whether it is possible to make a statement like (4.4.1) in an elegant and useful way (i.e., with manageable functions $f_{jii'}$ and $F_{jii'}$) and with a given joint confidence coefficient $1-\alpha$, i.e., free of the nuisance parameters Σ, is still an open question. It may well be that a range (not too wide) for the confidence coefficient itself is called for. Furthermore, whatever set of confidence intervals like (2.3.9.1) we propose, be it under a fixed confidence coefficient or under a confidence co-efficient lying in a short range, the 'goodness' of such a set would be a further question to tackle. For several reasons which need not be discussed here, it seems to the author that in this situation the more promising approach might be suitable two-stage procedures.

§ 4.5. General linear hypothesis and linear estimation. In place of the setup of subsection §4.2 , let us consider the following more general setup. Suppose we have a matrix $X(p \times n)$, consisting of n independently distributed p-dimensional column vectors $\underline{x}_1, \ldots, \underline{x}_n$, each being a multinormal with the same covariance matrix Σ. Suppose, further, that $E(X) = \xi(p \times m)\, B(m \times n)$ $(m < n)$, where B is a given (non-stochastic) matrix of rank $n_0 \leq m$ and $\xi(p \times m)$ is a set of unknown parameters. Suppose now that under this model we are interested in the problem of multiple or simultaneous estimation of a set of estimable linear vector parameters $\xi(p \times m)\, \underline{\ell}(m \times 1)$, for all $\underline{\ell}$ forming a vector space of rank $r \leq n_0 \leq m < n$. Also let $\underline{x}_{B,\underline{\ell}} \equiv X(p \times n)\, \underline{c}\ (n \times 1)$, be the best linear estimate of $\xi\underline{\ell}$ (notice that $\underline{c}$ can be obtained in terms of B and $\underline{\ell}$ and the estimate of the covariance matrix of $\underline{x}_{B,\underline{\ell}}$ to be called $S_{B,\underline{\ell}}$, is also available in terms of B and $\underline{\ell}$ and the $p \times n$ matrix of observations X). Thus, given B of rank $n_0 \leq m < n$, we have, for all non-null p-column-vectors $\underline{a}$ and all estimable linear functions $\xi\underline{\ell}$ (such that $\underline{\ell}$'s generate a vector space of rank $r \leq n_0$), by using the techniques of the previous sections, the set of simultaneous confidence interval statements (with confidence coefficient $1-\alpha$):

$$(4.5.1)\quad \underline{a}'\underline{x}_{B,\underline{\ell}} - [r\theta_\alpha(p,r,n-n_0)\underline{a}'S_{B,\underline{\ell}}\underline{a}]^{1/2} \leq \underline{a}'\xi\underline{\ell} \leq \underline{a}'\underline{x}_{B,\underline{\ell}} +$$

$$\left[r\theta_\alpha(p,r,n-n_0)\underline{a}'S_{B,\xi}\underline{a} \right]^{1/2} ,$$

where $\theta_\alpha(p,r,n-n_0)$ is defined in terms of the relevant parameters exactly the same way as in subsection §4.2 . The tie-up of (4.5.1) with the univariate confidence bounds given in (2.1.6) of § 2.1. will be obvious.

The inverse problem of testing of hypothesis would go through exactly the same way as in subsection §4.2 and need not be separately considered here.

5. Multivariate estimation and testing problems on covariance matrices.

§ 5.1. Problem on Σ from an $N(\underline{\xi}, \Sigma)$. As suggested in section 3, let us start from a $Y(p \times n)$: $N(\underline{0}, \Sigma)$, where $\Sigma(p \times p)$ is supposed to be p.d. (so that its characteristic roots are all positive) and where for simplicity we also assume that $p \leq n$, so that, a. e., YY', i.e., nS is p.d., so that a. e., all its characteristic roots are positive. We now recall the well-known result that there exists an orthogonal $\Gamma(p \times p)$ such that $\Sigma(p \times p) = \Gamma(p \times p)\, D_{\theta}(p \times p)\, \Gamma'(p \times p)$ where θ's are the characteristic roots of Σ. If the roots are distinct then by a convention, say taking all the elements of the first row of Γ to be positive, the transformation could be made one-to-one. But this we shall not need for our present purpose. Note that the number of independent elements on both sides is the same. We shall discuss the estimation and testing problems not in terms of Σ but in terms the equivalent set Γ and θ. Except for the factor $(-1/2)$ the argument under the exponential in the probability density of Y can now be written as

$\mathrm{tr}(\Gamma D_{\Theta} \Gamma')^{-1} YY' = \mathrm{tr}\, \Gamma D_{1/\sqrt{\Theta}} D_{1/\sqrt{\Theta}} \Gamma' YY' = \mathrm{tr}(D_{1/\sqrt{\Theta}} \Gamma' Y)(D_{1/\sqrt{\Theta}} \Gamma' Y)'$

If we put $Z = D_{1/\sqrt{\Theta}} \Gamma' Y$, it is easy to chedk that the probability density of Z is

$$(5.1.1) \qquad [\, 1/(2\pi)^{\frac{pn}{2}} \,] \operatorname{Exp} [\, -\tfrac{1}{2} \operatorname{tr} Z Z' \,]$$

Let us at this point try to obtain a set of simultaneous confidence bounds on a class of arbitrary p.d. quadratic functions of the elements of the population matrix $D_{1/\sqrt{\Theta}} \Gamma'$ (to be brought out in 5.1.5).

For all non-null non-stochastic $\underline{a}(p \times 1)$ consider now the simultaneous statement that

$$(5.1.2) \qquad c_1^2 \leq \underline{a}' ZZ' \underline{a}/\underline{a}'\underline{a} \leq c_2^2 \text{ or } c_1^2 \leq \underline{a}'(D_{1/\sqrt{\Theta}} \Gamma' YY' \Gamma D_{1/\sqrt{\Theta}})\underline{a}/\underline{a}'\underline{a} \leq c_2^2$$

This statement, for a given Z and c_1^2 and c_2^2 is precisely equivalent to the statement that

$$c_1^2 \leq \operatorname{Inf}_{\underline{a}} \frac{\underline{a}' ZZ' \underline{a}}{\underline{a}'\underline{a}} \leq \operatorname{Sup}_{\underline{a}} \frac{\underline{a}' ZZ' \underline{a}}{\underline{a}'\underline{a}} \leq c_2^2$$

(the "Sup" and "Inf" being with respect to variation over $\underline{a}$), or that

(5.1.3) $c_1^2 \leq \theta_1 \leq \theta_p \leq c_2^2$, where θ_1 and θ_p are the smallest and largest characteristic roots of the matrix ZZ', both, a.e., p.d. The relevant distributions on the null hypothesis, i.e., when the true population matrix is Σ, being known, let us determine c_1^2 and c_2^2 from the relations

$$(5.1.4) \qquad \begin{gathered} P(c_1^2 \leq \theta_1 \leq \theta_p \leq c_2^2 \mid \Sigma) = 1-\alpha \text{ and} \\ P(c_1^2 \leq \theta_1 \mid \Sigma) = P(\theta_p \leq c_2^2 \mid \Sigma) \end{gathered}$$

We can write c_1^2 and c_2^2 as $\theta_{1\alpha}(p,n)$ and $\theta_{2\alpha}(p,n)$.

If we now tie up (5.1.2), (5.1.3) and (5.1.4) we have, with a confidence coefficient $1-\alpha$, the set of multiple or simultaneous confidence interval statements for all non-null $\underline{a}$ and all permissible values of the unknown parameters Γ and Θ's:

$$(5.1.5) \qquad \underline{a}'\underline{a}\,\theta_{1\alpha}(p,n) \leq \underline{a}'(D_{1/\sqrt{\Theta}}\Gamma'YY'\Gamma D_{1/\sqrt{\Theta}})\underline{a} \leq \underline{a}'\underline{a}\theta_{2\alpha}(p,n)$$

or, remembering that $nS = YY'$,

$$\underline{a}'\underline{a}\theta_{1\alpha}(p,n) \leq \underline{a}'(D_{1/\sqrt{\Theta}}\, n\,\Gamma'S\Gamma D_{1/\sqrt{\Theta}})\underline{a} \leq \underline{a}'\underline{a}\theta_{2\alpha}(p,n) .$$

The confidence statement is on the parametric matrix $D_{1/\sqrt{\Theta}}\Gamma'$ which, as will be presently seen, plays the same part as σ in univariate problems. Furthermore, we note that (5.1.5) gives a set of simultaneous confidence bounds on a class of arbitrary p.d. quadratic functions of the elements of the population matrix $D_{1/\sqrt{\Theta}}\Gamma'$ such that the elements of the observed sample covariance matrix S also enter into the coefficients of the quadratic functions. Note that when $p = 1$, i.e., in the univariate case, $\Gamma = \Gamma' = 1$ (with the convention we are using), $\Sigma = \sigma^2$, $D_{1/\sqrt{\Theta}} = 1/\sigma$, $\underline{a}' = \underline{a}$ = a scalar, so that (5.1.5) will reduce to

$$(5.1.6)\qquad \chi^2_{1\alpha} \le ns^2/\sigma^2 \le \chi^2_{2\alpha} \text{ or } ns^2/\chi^2_{1\alpha} \ge \sigma^2 \ge ns^2/\chi^2_{2\alpha}$$

where $_{1\alpha}$ and $_{2\alpha}$ are just the lower and upper a/2-points of ^{2} with n D.F.

It is easy to see by inversion of (5.1.5) that for the associated hypothesis H_0: $\Sigma = \Sigma_0 = \Gamma_0 D_{\Theta_0} \Gamma_0'$ (say), we have the critical region:

$$(5.1.7)\qquad \phi_p \ge \theta_{2\alpha}(p,n) \text{ and/or } \phi_1 \le \theta_{1\alpha}(p,n),$$

where ϕ_p and ϕ_1 are the largest and smallest characteristic roots of the matrix $D_{1/\sqrt{\Theta_0}} \Gamma_0' YY' \Gamma_0 D_{1/\sqrt{\Theta_0}}$. The shortness of the confidence bounds (5.1.5) is tied up with the power of (5.1.7) and the general nature and properties of this have been already indicated in a previous paper [1].

§ 5.2. Problem of comparison between Σ_1 and Σ_2 from $N(\underline{\xi}_1, \Sigma_1)$ and $N(\underline{\xi}_2, \Sigma_2)$.

Let us start from $Y_i(p \times n_i)$: $N(\underline{0}, \Sigma_i)(i = 1,2)$, where we assume that $p \le n_1, n_2$, and that Σ_1 and Σ_2 are both p.d. so that the characteristic roots of $\Sigma_1 \Sigma_2^{-1}$ are all positive and those of $Y_1Y_1'(Y_2Y_2')^{-1}$, i.e., of $(n_1/n_2)\, S_1S_2^{-1}$ are, a. e., all positive. We recall that there exists a non-singular $\mu(p \times p)$ such that $\Sigma_1 = \mu D_{\Theta} \mu'$ and $\Sigma_2 = \mu\mu'$, where Θ's are the characteristic roots of $\Sigma_1\Sigma_2^{-1}$. If these roots are distinct, then by a convention, say taking all the elements of the first row of μ to be positive, the transformation could be made one-to-one. This we shall not need, but noting that the number of independent elements on both sides is the same we shall work in terms of μ and Θ 's, and not Σ_1 and Σ_2. Except for the factor (-1/2) the argument under the exponential in the probability density of Y_1 and Y_2 can be written as

$$(5.2.1) \qquad \operatorname{tr} [(\mu D_{\Theta} \mu')^{-1} Y_1 Y_1' + (\mu\mu')^{-1} Y_2 Y_2']$$

$$= \operatorname{tr} [(D_{1/\sqrt{\Theta}} \mu^{-1} Y_1)(D_{1/\sqrt{\Theta}} \mu^{-1} Y_1)' + (\mu^{-1} Y_2)(\mu^{-1} Y_2)']$$

If we now put $Z_1 = D_{1/\sqrt{\Theta}} \mu^{-1} Y_1$ and $Z_2 = \mu^{-1} Y_2$, it is easy to check that the probability density of Z_1 and Z_2 is

$$[1/(2\pi)^{\frac{p(n_1+n_2)}{2}}] \operatorname{Exp} [- \tfrac{1}{2} \operatorname{tr} (Z_1 Z_1' + Z_2 Z_2'] .$$

Let us try to obtain a set of simultaneous confidence bounds on a class of arbitrary p.d. quadratic functions of the elements of the population matrix $\mu D_{1/\sqrt{\Theta}} \mu^{-1}$ (to be brought out in 5.2.4). For all non-null non-stochastic $\underline{a}$ (p x 1) consider now the set of statements that

$$(5.2.2) \qquad c_1^2 \leq \underline{a}' Z_1 Z_1' \underline{a} \,/\, \underline{a}' Z_2 Z_2' \underline{a} \leq c_2^2 \quad \text{or}$$

$$c_1^2 \leq \underline{a}'(D_{1/\sqrt{\Theta}} \mu^{-1} Y_1)(D_{1/\sqrt{\Theta}} \mu^{-1} Y_1)' \underline{a} / \underline{a}'(\mu^{-1} Y_2)(\mu^{-1} Y_2)' \underline{a} \leq c_2^2$$

$$\text{or } \frac{n_2}{n_1} c_1^2 \leq \underline{a}'(D_{1/\sqrt{\Theta}} \mu^{-1} S_1 \mu'^{-1} D_{1/\sqrt{\Theta}}) \underline{a} / \underline{a}'(\mu^{-1} S_2 \mu'^{-1}) \underline{a} \leq \frac{n_2}{n_1} c_2^2$$

For a given Z_1, Z_2, c_1^2 and c_2^2 this statement is precisely equivalent to the statement that

$$c_1^2 \leq \mathrm{Inf}_{\underline{a}} \frac{\underline{a}'Z_1Z_1'\underline{a}}{\underline{a}'Z_2Z_2'\underline{a}} \leq \mathrm{Sup}_{\underline{a}} \frac{\underline{a}'Z_1Z_1'\underline{a}}{\underline{a}'Z_2Z_2'\underline{a}} \leq c_2^2$$

(the 'Sup' and 'Inf' bring with respect to variation over $\underline{a}$), or that

$$(5.2.3) \qquad c_1^2 \leq \theta_1 \leq \theta_p \leq c_2^2$$

where θ_1 and θ_p are the smallest and largest characteristic roots of the matrix $(Z_1Z_1')(Z_2Z_2')^{-1}$, both, a. e., positive. The relevant distributions on the null hypothesis {i. e., in this case, when the true population matrices are Σ_1 and Σ_2} being known, let us determine c_1^2 and c_2^2 from the relations formally similar to (5.1.4) and write c_1^2 and c_2^2 as $\theta_{1\alpha}(p,n_1,n_2)$ and $\theta_{2\alpha}(p,n_1,n_2)$, remembering that these $\theta_{1\alpha}$ and $\theta_{2\alpha}$ will be different in form from those given in (5.1.4). If we now tie up (5.2.2) and (5.2.3) and put $\underline{a}'\mu^{-1} = \underline{b}'$, we have (with a confidence coefficient $1-\alpha$), the set of simultaneous confidence interval statements for all non-null $\underline{b}$ and all permissible values of the unknown parameters μ and Θ 's:

$$(5.2.4) \quad \frac{n_2}{n_1}\,\theta_{1\alpha}\,(p,n_1,n_2)\underline{b}'S_2\underline{b} \leq \underline{b}'(\mu D_{1/\sqrt{\Theta}}\,\mu^{-1}S_1\mu'^{-1}D_{1/\sqrt{\Theta}}\,\mu')\,\underline{b}$$
$$\leq \frac{n_2}{n_1}\,\theta_{2\alpha}\,(p,\, n_1,\, n_2)\,\underline{b}'\,S_2\underline{b}$$

The confidence statement relates to the parametric matrix $\mu\, D_{1/\sqrt{\Theta}}\, u^{-1}$ which, as will be presently noticed, plays the same part as σ_1/σ_2 in univariate problems. It may be observed that (5.2.4) gives a set of confidence bounds on a class of arbitrary p.d. quadratic functions of the elements of the population matrix $\mu \cdot D_{1/\sqrt{\Theta}} \mu^{-1}$ such that the elements of the observed sample matrix S_1 also enter with the coefficients of the quadratic functions. As in the previous case, note that when p = 1, $\underline{b} = \underline{b}'$ = a scalar, $\Sigma_1 = \sigma_1^2$, $\Sigma_2 = \sigma_2^2$ (both scalars), $D_{1/\sqrt{\Theta}} = \sigma_2/\sigma_1$ and $\mu D_{1/\sqrt{\Theta}} \mu^{-1} = \sigma_2/\sigma_1$, so that (4.2.4) reduces to

$$(5.2.5) \qquad F_{1\alpha} \cdot s_1^2/s_2^2 \geq \sigma_1^2/\sigma_2^2 \geq F_{2\alpha} \cdot s_1^2/s_2^2$$

where $F_{1\alpha}$ and $F_{2\alpha}$ are just the lower and upper $\alpha/2$-points of the F-distribution with D. F. n_1 and n_2.

It is easy to see by inversion of (5.2.4) that, for the associated hypothesis $H_0 : \Sigma_1 = \Sigma_2$ which turns up if and only if $D_\theta = I(p)$, we have the critical region obtained in the previous paper [1], namely,

$$(5.2.6) \qquad \phi_p \geq \theta_{2\alpha}\ (p, n_1, n_2) \text{ and/or } \phi_1 \leq \theta_{1\alpha}\ (p, n_1, n_2);$$

where ϕ_p and ϕ_1 are the largest and smallest characteristic roots of the matrix

$$(\mu^{-1} Y_1 Y_1' \mu'^{-1})(\mu^{-1} Y_2 Y_2' \mu'^{-1})^{-1} \text{ i.e. of } \mu^{-1}(Y_1 Y_1')(Y_2 Y_2')^{-1} \mu, \text{ i. e.}$$

of $(Y_1 Y_1')(Y_2 Y_2')^{-1}$ i.e., of $(n_1/n_2)\ S_1 S_2^{-1}$.

The shortness of the confidence bounds (5.2.4) is tied up with the power of (4.1.6) and the properties of this power have been already discussed in the previous paper [1].

6. Multivariate estimation and testing problems on 'association' parameters.

§ 6.1. Problem on the regression coefficient in a bivariate normal population. Let two variates x_1 and x_2 be supposed to be distributed as a bivariate normal with variances σ_1^2 and σ_2^2 and correlation coefficient ρ, and let the sample variances (on a sample of size n+1) be denoted by s_1^2 and s_2^2, and the sample correlation coefficient by r. Also let $b_1 = s_1 r/s_2$ and $\beta_1 = \sigma_1\rho/\sigma_2$. It is easy to check that the variates $(x_1 - \beta_1 x_2)$ and x_2 are uncorrelated (in the population), so that when the population parameters are σ_1, σ_2 and ρ, $\sqrt{n-1}\, r^*/\sqrt{1-r^{*2}}$ is well known to have the t-distribution with (n-1) D. F., where r^* stands for the sample correlation between $(x_1 - \beta_1 x_2)$ and x_2, i.e.,

$$(6.1.1)\qquad r^* = (s_1 s_2 r - \beta_1 s_2^2)/(s_1^2 - 2\beta_1 s_1 s_2 r + \beta_1^2 s_2^2)^{1/2}\, s_2$$

$$= (s_1 r - \beta_1 s_2)/\left[(s_1 r - \beta_1 s_2)^2 + (1 - r^2)s_1^2\right]^{1/2}$$

$$= (b_1 - \beta_1)/\left[(b_1 - \beta_1)^2 + (1 - r)^2 s_1^2/s_2^2\right]^{1/2},$$

and, therefore,

$$(6.1.2)\qquad r^*/\sqrt{1 - r^{*2}} = (b_1 - \beta_1)/(1 - r^2)^{1/2}\,\frac{s_1}{s_2}\ .$$

Now consider the statement

$$(6.1.3)\qquad -t_\alpha(n - 1) \leq \sqrt{n - 1}\, r^*/\sqrt{1 - r^{*2}} \leq t_\alpha(n - 1),$$

where $t_\alpha(n - 1)$ gives the upper $\alpha/2$-point of the t-distribution with (n-1) D.F. This is easily seen to reduce to the following confidence statement on β_1 (with a confidence coefficient $1 - \alpha$):

(6.1.4)

$$b_1 - \frac{t_\alpha(n-1)}{\sqrt{n-1}}(1-r^2)^{1/2}\ \frac{s_1}{s_2} \leq \beta_1 \leq b_1 + \frac{t_\alpha(n-1)}{\sqrt{n-1}}(1-r^2)^{1/2}\ \frac{s_1}{s_2}$$

By inversion of (6.1.4) the test that we obtain for the associated hypothesis H_0: $\beta_1 = 0$, i.e., $\rho = 0$, is easily checked to be the customary test based on 'r' or just the t-test, and nothing further need be said about the power of this test, (which is well-known) or the shortness of the interval (6.1.4). Similar procedures would go through for 'partial regressions' or 'multiple regressions'. The interesting point to note here is that it would be far more difficult to give corresponding confidence bounds to ρ, because we shall have to do it by inverting the distribution of the non-central r, which is quite complicated.

§ 6.2. Problem on the regression like parameters in a (p + q)-variate normal population. Let us start from an $Y((p+q) \times n) : N(\underline{0}, \Sigma)$, where $p \leq q$, $p + q \leq n$ and where Σ is p.d. and of the form, say,

$$\begin{array}{cc} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} & \begin{matrix} p \\ q \end{matrix} \\ \begin{matrix} p \quad & \quad q \end{matrix} & \end{array}$$

, so that Σ_{11} and Σ_{22} themselves are also p.d.

In this case, all the p population canonical correlations, i.e., all characteristic roots θ_i's of $\Sigma_{11}^{-1}\ \Sigma_{12}\ \ \Sigma_{22}^{-1}\ \Sigma_{12}'$ are non-negative and less than 1. If now, Σ_{12} is of rank $s(\leq p \leq q)$, then s of these roots are positive and the rest, i. e., p-s are zero. We use now the theorem that there exist non-singular $\mu_1(p \times p)$ and $\mu_2(q \times q)$ such that

$$\Sigma_{11} = \mu_1\mu_1' \;;\; \Sigma_{22} = \mu_2\mu_2' \quad \text{and}$$

$$\Sigma_{12}(p \times q) = \mu_1(p \times p)\begin{array}{cc} (D_{\sqrt{\Theta}} & 0) \\ p & q-p \end{array}(p)\, \mu_2'(q \times q),$$

If Σ_{12} is of rank p and the Θ_i's (now all positive) are distinct, then this transformation could be made one-to-one by taking

$$\mu_2(q \times q) = \begin{array}{cc} p & q-p \\ \left(\begin{array}{c} \mu_{21} \\ \\ \mu_{23} \end{array}\right. & \left.\begin{array}{c} \tilde{\mu}_{22} \\ \\ \mu_{24} \end{array}\right) \end{array} \begin{array}{c} \\ q-p \\ \\ p \end{array}$$

and adopting the convention, say, that the elements of the first row of μ_1 and the diagonal elements of $\tilde{\mu}_{22}$ are all to be positive. If Σ_{12} is of rank $s(<p)$ and the s positive Θ_i's are distinct, then this transformation could be made unique by taking

$$D_{\Theta} = \begin{array}{c} \\ s \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cc} s & p-s \end{array} \\ \left(\begin{array}{c|c} \begin{matrix} \Theta_1 & & 0 \\ & \ddots & \\ 0 & & \Theta_s \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right) \end{array} \;;\; \mu_1(p \times p) = \begin{array}{cc} s & p-s \\ \left(\begin{array}{c} \mu_{11} \\ \mu_{23} \end{array}\right. & \left.\begin{array}{c} \tilde{\mu}_{12} \\ \mu_{24} \end{array}\right) \end{array} \begin{array}{c} \\ p-s \\ s \end{array} \;;$$

$$\mu_2(q \times q) = \begin{array}{cc} s & q-s \\ \left(\begin{array}{c} \mu_{21} \\ \\ \mu_{23} \end{array}\right. & \left.\begin{array}{c} \tilde{\mu}_{22} \\ \\ \mu_{24} \end{array}\right) \end{array} \begin{array}{c} \\ q-s \\ \\ s \end{array} \quad ,$$

and adopting the convention, say, that the first row of μ_{11} and the diagonals of $\tilde{\mu}_{12}$ and $\tilde{\mu}_{22}$ are all positive. We shall not need this uniqueness, but we note that with proper forms for μ_1 and μ_2 the number of independent elements is the same on both sides and we shall work in terms of μ_1, μ_2 and Θ 's and not the Σ. We now put

$$
Y = \begin{array}{c} n \\ \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \end{array} \begin{array}{c} \\ p \\ q \end{array}, \text{ so that } YY' = \begin{array}{cc} \quad p & \quad q \\ \multicolumn{2}{c}{\begin{pmatrix} Y_1Y_1' & Y_1Y_2' \\ Y_2Y_1' & Y_2Y_2' \end{pmatrix}} \end{array} \begin{array}{c} \\ p \\ q \end{array} = n \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix}
$$

We next observe that, a. e., YY' is p.d. (which means that S_{11} and S_{22} are p.d.) and S_{12} is of rank p, so that, a. e., all the p characteristic roots of $S_{11}^{-1}\ S_{12}\ S_{22}^{-1}\ S_{12}'$ are > 0 and < 1. We next note that

$$
\Sigma = \begin{pmatrix} \mu_1\mu_1' & \mu_1 \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \mu_2' \\ \mu_2 \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} \mu_1' & \mu_2\mu_2' \end{pmatrix}
$$

$$
= \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \begin{pmatrix} I(P) & 0 \\ \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} & \begin{pmatrix} D_{\sqrt{1-\Theta}} & 0 \\ 0 & I(q-p) \end{pmatrix} \end{pmatrix} \begin{pmatrix} I(p) & \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \\ 0 & \begin{pmatrix} D_{\sqrt{1-\Theta}} & 0 \\ 0 & I(q-p) \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mu_1' & 0 \\ 0 & \mu_2' \end{pmatrix},
$$

so that

$$
\Sigma^{-1} = \begin{pmatrix} \mu_1'^{-1} & 0 \\ 0 & \mu_2'^{-1} \end{pmatrix} \begin{pmatrix} I(p) & -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} & 0 \end{pmatrix} \\ 0 & \begin{pmatrix} D_{\sqrt{1/1-\Theta}} & 0 \\ 0 & I(q-p) \end{pmatrix} \end{pmatrix} \begin{pmatrix} I(p) & 0 \\ -\begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} \\ 0 \end{pmatrix} & \begin{pmatrix} D_{\sqrt{1/1-\Theta}} & 0 \\ 0 & I(q-p) \end{pmatrix} \end{pmatrix}
$$

$$x \begin{pmatrix} \mu_1^{-1} & 0 \\ 0 & \mu_2^{-1} \end{pmatrix}$$

Except for the factor (-1/2), the argument under the exponential in the probability density of Y_1 and Y_2 can be now written as

$$\mathrm{tr} \left[\begin{matrix} \mu_1^{-1} Y_1 \\ -\begin{pmatrix} {}^{D}\sqrt{\theta/1-\theta} \\ 0 \end{pmatrix} \mu_1^{-1} Y_1 + \begin{pmatrix} {}^{D}\sqrt{1/1-\theta} & 0 \\ 0 & I \end{pmatrix} \mu_2^{-1} Y_2 \end{matrix} \right] \left[\quad \right]$$

If we now put

(6.2.1) $$Z_1 = \mu_1^{-1} Y_1 \text{ and } Z_2 = - \begin{pmatrix} {}^{D}\sqrt{\theta/1-\theta} \\ 0 \end{pmatrix} \mu_1^{-1} Y_1 + \begin{pmatrix} {}^{D}\sqrt{1/1-\theta} & 0 \\ 0 & I \end{pmatrix} u_2^{-1} Y_2$$

$$= - \delta_1 \mu_1^{-1} Y_1 + \delta_2 \mu_2^{-1} Y_2 \text{ (say, for shortness)},$$

it is easy to check that the probability density of Z_1 and Z_2 is

$$\left[1/(2\pi)^{\frac{(p+q)n}{2}} \right] \mathrm{Exp} \left[-\tfrac{1}{2} \mathrm{tr} \, (Z_1 Z_1' + Z_2 Z_2') \right]$$

Here we shall be interested in a set of simultaneous confidence bounds on a certain class of arbitrary p.d. quadratic functions (to be brought out in (6.2.9) of the elements of the population matrix $\mu_1'^{-1}$(p x p) $({}^{D}\sqrt{\rho_{i}}_{\,p} \;\; 0_{\,q-p})$ (p) u_2' (q x q). For all pairs of non-null and non-stochastic $\underline{a}_1$ (p x 1) and $\underline{a}_2$ (q x 1), consider now the set of statements that

(6.2.2) $$(a_1' Z_1 Z_2' \underline{a}_2)^2 / (\underline{a}_1' Z_1 Z_1' \underline{a}_1)(\underline{a}_2' Z_2 Z_2' \underline{a}_2) \leq c^2$$

For a given Z_1, Z_2 and c^2 this is precisely equivalent to the statement that

$$\operatorname{Sup}_{\underline{a}_1, \underline{a}_2} \; (\underline{a}_1' Z_1 Z_2' \underline{a}_2)^2 / (a_1' Z_1 Z_1' a_1)(a_2' Z_2 Z_2' a_2) \leq c^2$$

(the sup being with respect to variation over $\underline{a}_1$ and $\underline{a}_2$) or that

(6.2.3) $\quad \Theta_p \leq c^2,$

where Θ_p is the largest (and of course positive) characteristic root of

$(Z_1Z_1')^{-1}(Z_1Z_2')(Z_2Z_2')^{-1}(Z_2Z_1')$. The relevant distribution on the null hypothesis, i.e., when the true population matrix is Σ, being known, let us determine c^2 from the relation:

$P(\Theta_p \leq c^2 \mid \text{true population matrix} = \Sigma) = 1 - \alpha$, and then write c^2

as Θ_α or $\Theta_\alpha(p,q,n)$. Next note that, with

(6.2.4) $$\delta_1 = \begin{array}{cc} \begin{pmatrix} D\sqrt{\Theta/1-\Theta} \\ 0 \end{pmatrix} & \begin{array}{c} p \\ q-p \end{array} \\ p & \end{array} \quad \text{and} \quad \delta_2 = \begin{array}{cc} \begin{pmatrix} D\sqrt{1/1-\Theta} & 0 \\ 0 & I(q-p) \end{pmatrix} & \begin{array}{c} p \\ q-p \end{array} \\ \begin{array}{cc} p & q-p \end{array} & \end{array} = \delta_2',$$

we have from (6.2.1)

(6.2.5) $$Z_1Z_1' = n\mu_1^{-1} S_{11}\mu_1'^{-1}; \quad Z_1Z_2' = n\mu_1^{-1} [-S_{11}\mu_1'^{-1}\delta_1' + S_{12}\mu_2'^{-1}\delta_2]$$

$$Z_2Z_2' = n[\delta_1\mu_1^{-1}S_{11}\mu_1'^{-1}\delta_1' - \delta_1'\mu_1^{-1}S_{12}\mu_2'^{-1}\delta_2 - \delta_2\mu_2^{-1}S_{12}'\mu_1'^{-1}\delta_1' + \delta_2\mu_2^{-1}S_{22}\mu_2'^{-1}\delta_2]$$

If we now put $\underline{a}_1'\mu_1^{-1} = \underline{b}_1'$ and $\underline{a}_2' \; \delta_2\mu_2^{-1} = \underline{a}_2' \begin{array}{c} \begin{pmatrix} D\sqrt{1/1-\Theta} & 0 \\ 0 & I \end{pmatrix} \\ \begin{array}{cc} p & q-p \end{array} \end{array} \begin{pmatrix} p \\ q-p \end{pmatrix} \mu_2^{-1} = \underline{b}_2',$

and

tie up all relations from (6.2.2) to (6.2.5), we have for all non-null $\underline{a}_1$ and $\underline{a}_2$ and all permissible μ_1, μ_2 and Θ s the following set of simultaneous confidence interval statements (with a confidence coefficient $1 - \alpha$):

$$\text{(6.2.6)} \qquad \frac{\left[\underline{b}_1'(-S_{11}\mu_1'^{-1}\delta_1'\delta_2^{-1}\mu_2' + S_{12})\,\underline{b}_2\right]^2}{(\underline{b}_1'S_{11}\underline{b}_1)\left[\underline{b}_2'(\mu_2\delta_2^{-1}\delta_1\mu_1^{-1}S_{11}\mu_1'^{-1}\delta_1'\delta_2^{-1}\mu_2' - \mu_2\delta_2^{-1}\delta_1\mu_1^{-1}S_{12} - (\ )' + S_{22})\,\underline{b}_2\right]} \leq \theta_\alpha(p,\ q,\ n)$$

Note that

$$\text{(6.2.7)} \qquad \delta_1'\delta_2^{-1} = p\begin{array}{c} \begin{pmatrix} D_{\sqrt{\Theta/1-\Theta}} & 0 \end{pmatrix} \\ \begin{array}{cc} p & q-p \end{array} \end{array} \begin{array}{c} \begin{pmatrix} D_{\sqrt{1-\Theta}} & 0 \\ 0 & I \end{pmatrix} \\ \begin{array}{cc} p & q-p \end{array} \end{array} \begin{array}{c} p \\ q-p \end{array} = \begin{array}{c} \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \\ \begin{array}{cc} p & q-p \end{array} \end{array} p\,,$$

so that putting

$$\text{(6.2.8)} \qquad \beta(p \times q) = \mu_1'^{-1}\,(p \times p)\ \begin{array}{c} \begin{array}{cc} p & q-p \end{array} \\ \begin{pmatrix} D_{\sqrt{\Theta}} & 0 \end{pmatrix} \end{array}\,(p)\ \mu_2'(q \times q)$$

we have, for this β, the set of confidence statements

$$\text{(6.2.9)} \qquad \frac{\left[\underline{b}_1'\,(-S_{11}\,\beta + S_{12})\,\underline{b}_2\right]^2}{(\underline{b}_1'S_{11}\underline{b}_1)\left[\underline{b}_2'(\beta'S_{11}\beta - \beta'S_{12} - S_{12}'\,\beta + \beta_{22})\,\underline{b}_2\right]} \leq \theta_\alpha(p,q,n)$$

(6.2.9) gives a set of simultaneous confidence bounds on a class of arbitrary p. d. quadratic functions of the elements of the population matrix β such that the elements of the observed sample matrices

$$S_{11},\ S_{22} \text{ and } S_{12}$$

also enter into the coefficients of the class of arbitrary functions. It is interesting to observe that when $p = q = 1$, we may take $\mu_1 = \mu_1' = \sigma_2$, $\mu_2 = \mu_2' = \sigma_1$ and $D_{\sqrt{\Theta}} = \rho$, so that $\beta = \sigma_1\rho/\sigma_2$ and check that (6.2.9) reduces to (6.1.4) for the regression coefficient. Indeed the β given by (6.2.8) can really be regarded as the regression of the set of p variates on the set of q variates or in other words, an appropriate generalization of bivariate regression coefficient.

It is easy to check by inversion of (6.2.9) that for the associated hypothesis H_0: $\beta = 0$, i.e., $D_{\sqrt{\Theta}} = 0$, i.e., $\Sigma_{12} = 0$, we have the critical region obtained in the previous paper [1], namely

$$\phi_p \geq \Theta_\alpha(p,q,n), \tag{6.2.10}$$

when ϕ_p is the largest characteristic root of the matrix $(Y_1Y_1')^{-1}(Y_1Y_2')(Y_2Y_2')^{-1}(Y_2Y_1')$, i.e., of the matrix $S_{11}^{-1}S_{12}S_{22}^{-1}S_{12}'$.

The shortness of the confidence bounds (6.2.9) is tied up with the power of (6.2.10) and the properties of this power have been already discussed in the previous paper [1].

References

[1] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," Annals of Math. Stat., Vol. 24(1953), pp.

[2] H. Scheffe, "A method for judging all contrasts in the analysis of variance," read at the East Lansing meeting of the Institute of Mathematical Statistics, September, 1952, mimeographed paper.

[3] J. W. Tukey, "Allowances for various types of error rates," unpublished invited address, Blacksburg meeting of the Institute of Mathematical Statistics, March, 1952.