ON A HEURISTIC METHOD OF TEST CONSTRUCTION AND ITS USE IN MULTIVARIATE ANALYSIS

By S. N. Roy

University of North Carolina

1. Summary. In this paper two closely related heuristic principles of test construction (to be explained in Section 3), called Type I and Type II methods, of which Type II is identified with the usual likelihood ratio method, are noticed as underlying most of the classical tests of hypotheses, simple or composite, on means of univariate normal populations, and on total or partial correlations or regressions in the case of multinormal variates. In these situations the two methods are found to lead to identical tests having properties which happen to be very good in certain cases and moderately good in others. For certain types of composite hypotheses an extension is then made of the Type I method which is applied to construct tests of three different classes of hypotheses on multinormal populations (so as to cover, between them, a very large area of multivariate analysis), yielding in each case a test in general different from the corresponding and current likelihood ratio test. In each case, however, the two tests happen to come out identical for some degenerate "degrees of freedom." In contrast to the likelihood ratio test it is found that in these cases, for general "degrees of freedom," the corresponding Type I test is much easier to use on small samples, because of the relatively greater simplicity of the corresponding small sample distribution problem under the null hypothesis. In each case a lower bound of the power function of the Type I test is also given (against all relevant alternatives), anything like which, so far as the author is aware, would be far more difficult to obtain for the Type II tests in these situations. In this paper the *general* approach to the two methods is entirely of a heuristic nature except that, under fairly wide conditions, a lower bound to the power functions for each of the two types of tests is indicated to be readily available, which, however, is much too crude or wide a bound in *general*.

2. Notation and preliminaries. As far as possible observations and sample quantities will be noted by Roman letters and population parameters by Greek letters; scalars by small letters, matrices by capital letters, column vectors by small letters underscored, and row vectors by priming them; the determinant of a square matrix M by $| M |$; "positive definite" by p.d.; "positive semidefinite" by p.s.d.; "except for a set of points of probability measure zero" or "almost everywhere" by a.e. $A(:p \times q)$ will indicate that the matrix A is $p \times q$, and $I(p)$ will stand for a $p \times p$ unit matrix. $x:N(\xi, \sigma^2)$, $\underline{x}(:p \times 1):N(\underline{\xi}(:p \times 1), \Sigma(:p \times p))$ and $X(:p \times n):N(\underline{\xi}(:p \times 1), \Sigma(:p \times p))$ will indicate respectively that the scalar x is normally distributed about a mean ξ with a variance σ^2, the

Received 12/20/52.

column vector $\underset{\sim}{x}$ is multi-normally distributed about a mean vector $\underset{\sim}{\xi}$ with a (p.d.) covariance matrix Σ, and the n column vectors of the $p \times n$ matrix X are independently and multinormally distributed, each column about a mean vector $\underset{\sim}{\xi}$ with a (p.d.) covariance matrix Σ. Exceptions to this notation will be clearly indicated at the proper places. For the sake of clarity, it may be noted that the X above has the probability density

$$[1 \mid (2\pi)^{-\frac{1}{2}pn} \mid \Sigma \mid^{\frac{1}{2}n}] \exp [-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(X - \xi)(X' - \xi')],$$

where ξ $(:p \times n)$ is a $p \times n$ matrix each column of which is the column vector $\underset{\sim}{\xi}$. Furthermore, $d\underset{\sim}{x}$ will stand for $\prod_{i=1}^{p} dx_i$ and dX for $\prod_{\lambda=1}^{n} \prod_{i=1}^{p} dx_{i\lambda}$.

Throughout this paper all *general* discussions will be made in terms of the denumerable case, because I feel that perhaps the ideas are made clearest that way. The extension to the nondenumerable case might in *general* lead to measure theoretic difficulties, but such difficulties do not arise in the applications (most of them being nondenumerable cases) treated here.

The most powerful critical region of size, say $\beta_i(<1)$, which under fairly general conditions will exist and which under slightly less general conditions will also be unique), of a simple hypothesis H_0 against a simple alternative H_i (such that $H_0, H_i \varepsilon \Omega$, where $i = 1, 2, \cdots$, and where Ω stands for a domain of possibilities) will be denoted by $\omega(H_0, H_i, \beta_i)$, its complement, the acceptance region, by $\bar{\omega}(H_0, H_i, \beta_i)$, to indicate that in *general* both will depend upon H_0, H_i and β_i. The union of regions $\omega(H_0, H_i, \beta_i)$ over different H_i, β_i or $i(i = 1, 2, \cdots)$ will be denoted by $U_{H_i}\omega(H_0, H_i, \beta_i)$ or simply by $U_i\omega(H_0, H_i, \beta_i)$, and the intersection of regions $\bar{\omega}(H_0, H_i, \beta_i)$ over different $i(i = 1, 2, \cdots)$ by $\bigcap_{H_i}\bar{\omega}(H_0, H_i, \beta_i)$ or simply by $\bigcap_i \bar{\omega}(H_0, H_i, \beta_i)$. $P(H_0, H_i, \beta_i)$ will stand for the power of the most powerful test at level β_i or H_0 against H_i, and will in *general* depend upon all the three elements. It can be easily proved and has been published in an earlier paper [14] that $P(H_0, H_i, \beta_i) > \beta_i$. For convenience a sketch is given here. Assume, for simplicity of discussion but without any essential loss of generality, that we have a set of n continuous stochastic variates, $\underset{\sim}{x}(:n \times 1)$ or simply $\underset{\sim}{x}$, with respective probability densities $\phi_{H_0}(\underset{\sim}{x})$ and $\phi_{H_i}(\underset{\sim}{x})$ (or simply ϕ_{H_0} and ϕ_{H_i}) under the hypotheses H_0 and H_i. Then it is well known that $\omega(H_0, H_i, \beta_i)$ and $\bar{\omega}(H_0, H_i, \beta_i)$ are given respectively by

$$\begin{aligned} &\omega(H_0, H_i, \beta_i) : \phi_{H_i} \geqq \lambda\phi_{H_0}, \\ &\bar{\omega}(H_0, H_i, \beta_i) : \phi_{H_i} < \lambda\phi_{H_0}, \end{aligned} \tag{2.1}$$

where λ is determined by

$$P(\underset{\sim}{x} \; \varepsilon \; \omega(H_0, H_i, \beta_i) \mid H_0) : \beta_i.$$

Assume here that ϕ is such that ω defined by (2.1) is unique. Integrating the first inequality of (2.1) over $\omega(H_0, H_i, \beta_i)$ and the second one over $\bar{\omega}(H_0, H_i, \beta_i)$ we have respectively $P(H_0, H_i, \beta_i) \geqq \lambda\beta_i$ and $1 - P(H_0, H_i, \beta_i) < \lambda(1 - \beta_i)$, from which, after a slight reduction, we have

$$P(H_0, H_i, \beta_i) > \beta_i. \tag{2.2}$$

Note that in general λ will be of the form $\lambda(H_0, H_i, \beta_i)$, depending on all the elements. Incidentally, *any* critical region of size β for H_0, whose power with respect to an alternative H is greater than or equal to β, will be called an *unbiased critical region for H_0 against H.*

The likelihood ratio critical region at a level, say α, of H_0 against the whole class $H_i \varepsilon \Omega$, provided that it exists, will be denoted by $\hat{\omega}(H_0, \alpha)$. As is well known it is given by

$$\hat{\omega}(H_0, \alpha) : \phi(\underset{\sim}{x}) \geqq \mu(H_0, \alpha)\phi_{H_0}(\underset{\sim}{x}), \tag{2.3}$$

where, for a given $\underset{\sim}{x}$, $\phi(\underset{\sim}{x})$ stands for the largest $\phi_{H_i}(\underset{\sim}{x})$ (provided that it exists) with respect to variation of H_i over Ω, and where $\mu(H_0, \alpha)$ is given by

$$P(\underset{\sim}{x} \varepsilon \hat{\omega}(H_0, \alpha) \mid H_0 = \alpha. \tag{2.4}$$

Notice that $\phi(\underset{\sim}{x})$ is a function of $\underset{\sim}{x}$ only, being independent of H_i, but may depend on the *total* domain Ω. The power of this test, against any alternative H_i will be denoted by $P(H_0, H_i, \alpha)$.

Assume now that H_0 is a composite hypothesis and $H_i (i = 1, 2, \cdots)$ a composite alternative. In earlier papers [8], [13], [14] the author gave a set of sufficient conditions on ϕ_{H_0} for the availability of similar regions for H_0, and a set of (further) restrictions on ϕ_{H_i} and ϕ_{H_0} for the availability, among these similar regions, of one which is most powerful for H_0 against H_i in the following sense. Suppose H_0 and H_i are composite hypotheses, each characterized by some specified and some unspecified elements, so that, if the unspecified elements were specified, both H_0 and H_i would be simple hypotheses. Now suppose that, among the similar regions for H_0, there is one whose location in the sample space depends on the specified elements of H_0 and possibly on those of H_i, but not on the unspecified elements of H_0 or H_i, but which is nevertheless the most powerful critical region for *any* simple hypothesis within H_0 (obtained by specifying the unspecified elements) against *any* simple alternative within H_i (obtained by specifying the unspecified elements). But this "most powerful" is "most powerful among similar regions." If we drop the restriction of similarity and set up in a straightforward manner the most powerful critical region for the simple hypothesis in question against the simple alternative in question, then we may get a (*nonsimilar*) region having a larger power than that of the most powerful *similar* critical region just referred to. Such a most powerful similar critical region may be conveniently called a *bisimilar* region for H_0 against H_i. The likelihood ratio critical region for composite H_0 against all composite $H_i \varepsilon \Omega$ (which we know how to construct, provided that it exists), can be shown [13], [14] to be a similar region for H_0, under the restrictions just referred to. In this situation the same notation will be used as introduced in the previous paragraph for the case of a simple hypothesis against simple alternatives, and the result (2.2) will also hold, it being noted that, while the regions will be independent of the unspecified elements in H_0 and H_i, $P(H_0, H_i, \beta_i)$ and $P(H_0, H_i, \alpha)$ however, might depend on the unspecified elements of H_i though not on those of H_0.

3. Type I and Type II tests.

3.1. *Definitions and some remarks.* Consider, for simplicity of discussion but without any essential loss of generality (for the definitions could be immediately carried over into the case of composite hypothesis and alternative) a simple hypothesis H_0 against a simple alternative H_i such that H_0, $H_i(i = 1, 2, \cdots) \in \Omega$.

(i) Put $\beta_i = \beta(i = 1, 2, \cdots)$ and set up as the rejection and acceptance regions for $H_0 U_i\omega(H_0, H_i, \beta)$ and its complement $\bigcap_i\bar{\omega}(H_0, H_i, \beta)$ to be called, respectively, U_i and $\bigcap_i$. *This is defined to be a Type I test for H_0 against the whole class $H_i \varepsilon \Omega$*, the level of significance α being given by

$$P(\underline{x} \varepsilon U_i\omega(H_0, H_i, \beta) \mid H_0) = \alpha(H_0, \beta), (>\beta). \tag{3.1.1}$$

Let us for the moment assume nontriviality, that is, that given $\alpha < 1$, we can find $\beta = \beta(H_0, \alpha) > 0$, for which (3.1.1) will hold.

(ii) Put, in Section 2, $\lambda(H_0, H_i, \beta_i) = \mu$ (a preassigned constant) for all $i = 1, 2, \cdots$, and rewrite $\omega(H_0, H_i, \beta_i)$ and $\bar{\omega}(H_0, H_i, \beta_i)$ as $\omega^*(H_0, H_i, \mu)$ and $\bar{\omega}^*(H_0, H_i, \mu)$.

Now set up, as the rejection and acceptance regions for H_0, $U_i\omega^*(H_0, H_i, \mu)$ and its complement $\bigcap_i\bar{\omega}^*(H_0, H_i, \mu)$, to be called, respectively, U_i^* and $\bigcap_i^*$, where the β_i's ($i = 1, 2, \cdots$) are subject to $\lambda(H_0, H_i, \beta_i) = \mu$ (a preassigned constant). *This is defined to be a Type II test for H_0 against the whole class $H_i \varepsilon \Omega$* the level of significance α^* being given by

$$P(\underline{x} \varepsilon U_i^*(H_0, H_i, \mu) \mid H_0) = \alpha^*(H_0, \mu). \tag{3.1.2}$$

Here again let us, for the moment, assume nontriviality, that is, that given $\alpha^*(<1)$, we can find a μ such that $\beta(H_0, H_i, \mu) = \beta_i(>0)$ and that (3.1.2) will hold. This can be easily recognized as the likelihood ratio test by the following consideration. Notice that $\omega^*(H_0, H_i, \mu)$ (with a preassigned μ) is given by

$$\omega^*(H_0, H_i, \mu): \phi_{H_i}(\underline{x}) \geqq \mu\phi_{H_0}(\underline{x}). \tag{3.1.3}$$

Any $\underline{x}$ would belong to $U_i\omega^*(H_0, H_i, \mu)$ if for that $\underline{x}$, there were at least one $H_i \varepsilon \Omega$ for which (3.1.3) holds. It is easy to see that this would be accomplished if for that $\underline{x}$ the largest $\phi_{H_i}(\underline{x})$ (under variation of H_i over Ω) were $\geqq \mu\phi_{H_0}(\underline{x})$. Hence it is obvious that

$$\begin{aligned} &U_i\omega^*(H_0, H_i, \mu): \phi(\underline{x}) \geqq \mu\phi_{H_0}(\underline{x}) \\ &\bigcap_i\omega^*(H_0, H_i, \mu): \phi(\underline{x}) < \mu\phi_{H_0}(\underline{x}). \end{aligned} \tag{3.1.4}$$

3.2. *An obvious property of the two types of test.* Notice that U_i includes all $\omega(H_0, H_i, B)$ and U_i^* all $\omega^*(H_0, H_i, \mu)$. Now putting

$$P(\underline{x} \varepsilon U_i \mid H_i) \equiv P(U_i, H_i, \alpha) \text{ and } P(\underline{x} \varepsilon U_i^* \mid H_i) \equiv P(U_i^*, H_i, \alpha)$$

we shall have from Sections (2) and (3) for the two types of tests

$$\beta(H_0, \alpha) \equiv \beta < P(H_0, H_i, \beta) \leqq P(U_i, H_i, \alpha) \leqq P(H_0, H_i, \alpha) \leqq 1 \tag{3.2.1}$$
$$P(H_0, H_i, \alpha) > \alpha$$

$$\begin{aligned} \beta^*(H_0, H_i, \alpha) \equiv \beta_i^* &< P^*(H_0, H_i, \mu) \\ &\leqq P(U_i^*, H_i, \alpha) \leqq P(H_0, H_i, \alpha) \leqq 1 \end{aligned} \tag{3.2.2}$$
$$P(H_0, H_i, \alpha) > \alpha.$$

(3.2.1) and (3.2.2) give respectively, for all $H_i \varepsilon \Omega$, the lower bounds $P(H_0, H_i, \beta)$ and $P^*(H_0, H_i, \mu)$ for $P(U_i, H_i, \alpha)$ and $P(U_i^*, H_i, \mu)$, which, however, in general, would be far from close except sometimes for large "deviations" from H_0. With more knowledge of the forms of ϕ_{H_0} and ϕ_{H_i} it is often possible to get far closer lower bounds; even the actual powers are often computable without much difficulty (and turn out to be pretty high) as for example in most of the classical tests on normal populations.

It is easy to see that the results of (3.1) and (3.2) could be easily generalized to cover the case of composite H_0 against composite $H_i \varepsilon \Omega$ provided that we have similar regions for H_0 and a bisimilar region for H_0 against H_i. This, therefore, need not be separately treated.

3.3. *Display of two classical tests as Type I tests.* (i) Almost all classical tests on univariate and multivariate normal populations (ii) most classical tests on other types of populations and (iii) many tests on multivariate normal populations proposed in recent years are known to be derivable (and indeed many of them have, in fact, been derived) from the "likelihood ratio" principle, so that they belong to Type II. The author finds that all the customary tests in category (i), for example, the test of significance of (1) a mean, (2) a mean difference, (3) total or partial or multiple correlation, and (4) regressions, (5) the F-test in analysis of variance, (6) the test of the hypothesis of equality of standard deviations for two univariate normal populations, (7) the test based on Hotelling's T, all belong to Type I as well. Those classical tests in category (ii) that the author has examined so far also all belong to Type I. Coming to those situations that are sought to be handled by tests proposed under category (iii), the author finds that the likelihood ratio tests offered so far, while they automatically belong to Type II, do not belong to Type I. On the other hand, if, in these situations, one carries out (see Section 5) the spirit and method of discriminant analysis, one gets tests (see Section 6) which belong to Type I in a sense slightly more general than we have indicated so far.

In this section we consider, for illustration, two well known classical tests and show that they belong to Type I.

(i) For $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$ the classical test of $H(\xi_1 = \xi_2) \equiv H_0$ against $H(\xi_1 \neq \xi_2) \equiv H$ at a level α is based on a critical region given by

$$t \geqq t_0 \quad \text{or} \quad \leqq -t_0, \tag{3.3.1}$$

where

$$t \equiv (n_1 + n_2 - 2)^{\frac{1}{2}}\{n_1n_2/[n_1 + n_2]\}^{\frac{1}{2}}x(\bar{x}_1 - \bar{x}_2)/\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\}^{\frac{1}{2}},$$

and t_0 is given by $P(t \geqq t_0 \mid H_0) = \alpha/2$ and where $(\bar{x}_1, \bar{x}_2)$, (s_1, s_2) stand for the means and standard deviations of two random samples of sizes n_1 and n_2 drawn from $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$, respectively. This is well known as a likelihood ratio test but it is easily checked as Type I as well, in the following way. It is well known that $t \geqq t_0$ is a one-sided uniformly most powerful (bisimilar) region of size $\alpha/2$ for the composite H_0 against the composite $H(\xi_1 > \xi_2) \equiv H_1$ and so also is $t \leqq -t_0$ for H_0 against $H(\xi_1 < \xi_2) \equiv H_2$; taking the union we have (3.3.1) of size α.

(ii) Consider the testing of a general linear hypothesis in analysis of variance which, as is well known, can be *formally* reduced to the following. Suppose we have random samples of sizes n_i, means $\bar{x}_i$ and standard deviations s_i, drawn respectively from $N(\xi_i, \sigma^2)(i = 1, \cdots, k)$, and suppose we want to test $H(\xi_1 = \xi_2 = \cdots = \xi_k) \equiv H_0$ against the whole class H of $(\xi_1, \cdots, \xi_k)$ violating H_0. Put $n \equiv \sum_{i=1}^k n_i$; $\bar{x} \equiv \sum_{i=1}^k n_i\bar{x}_i/n$; $\xi \equiv \sum_{i=1}^k n_i\xi_i/n$. Now the classical F-test for H_0, which is well known to be a likelihood ratio or Type II test has at a level α the critical region given by

$$F \geqq F_0 \qquad (3.3.2)$$

where $F \equiv [\sum_{i=1}^k n_i(\bar{x}_i - \bar{x})^2/(k - 1)] \div [\sum_{i=1}^k (n_i - 1)s_i^2/(n - k)]$ and where F_0 is given by $P(F \geqq F_0 \mid H_0) = \alpha$.

To recognize this as a Type I test as well we proceed as follows. It was observed in earlier papers [8], [13] that among similar regions for H_0 (which exist) there is a most powerful (bisimilar) region for H_0 against any specific $(\xi_1, \cdots, \xi_k) \equiv \underline{\xi}$ violating H_0, the region of size, say, β being given by

$$t \geqq t_0 \qquad (3.3.3)$$

where $t \equiv \sqrt{n - 2} \cot \theta$;

$$\cos \theta \equiv \sum_{i=1}^k n_i(\bar{x}_i - \bar{x})(\xi_i - \xi) \Bigg/ \left[\sum_{i=1}^k \{n_i(\bar{x}_i - \bar{x})^2 + (n_i - 1)s_i^2\}\right]^{\frac{1}{2}} \cdot \left[\sum_{i=1}^k n_i(\xi_i - \xi)^2\right]^{\frac{1}{2}}$$

and where t_0 is given by $P(t \geqq t_0 \mid H_0) = \beta$. It was also noticed in those papers that this t has exactly the usual t-distribution with $(n - 2)$ degrees of freedom. Notice that $t_0 \equiv t_0(n, \beta)$ and $\beta \equiv \beta(n, t_0)$. To obtain now the union of regions: $t \geqq t_0$ over different sets of $(\xi_1, \cdots, \xi_k)$ we note that a given set of (observed) $\bar{x}_i$'s and s_i's would belong to the union, if for that set there were at least one t such that $t \geqq t_0$. The union is thus easily checked to be given by: the largest t (by varying over $\xi_1, \cdots, \xi_k) \geqq t_0$ (which is fixed). But by (3.3.3) the largest t

would correspond to the largest value of $\cos\theta$, and, given $\bar{x}_i$'s and s_i's, the largest value of $\cos\theta$ (under variation over $\xi_1, \cdots, \xi_k$) is easily seen to be given by:

$$(3.3.4)\qquad \cos\theta \equiv \left[\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right]^{\frac{1}{2}} \Big/ \left[\sum_{i=1}^{k} \{(n_i - 1)s_i^2 + n_i(\bar{x}_i - \bar{x})^2\}\right]^{\frac{1}{2}},$$

so that the largest t is given by

$$(3.3.5)\qquad t \equiv (n-2)^{\frac{1}{2}}\left[\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right]^{\frac{1}{2}} \Big/ \left[\sum_{i=1}^{k} (n_i - 1)s_i^2\right]^{\frac{1}{2}}.$$

Therefore the union of regions: $t \geqq t_0$, is given exactly by (3.3.2), which is the critical region of the F-test. Notice that given the α of the F-test, F_0 is obtained from (3.3.2) in the form $F_0(k-1, n-k; \alpha)$; and next by identifying the union of regions $t \geqq t_0$, with $F \geqq F_0$ we have

$$t_0 \equiv [(k-1)(n-2)F_0/(n-k)]^{\frac{1}{2}} \equiv t_0(k-1, n-k; \alpha);$$

and next from (3.3.3) we have

$$\beta \equiv \beta(n, t_0) \equiv \beta(k-1, n-k; \alpha).$$

3.4. *Some further remarks on the two types of test.* It may be noted (See Sections 2 and 3) that by specializing the β_i's (the sizes of the most powerful critical regions against different alternatives) in *two* special ways we get in a heuristic manner the two types of test. By specializing the β_i's in other ways other heuristic principles could be set up, some of which, in special situations, might be "better" than the Type I or Type II tests. It has already been observed that in many situations Type I and Type II tests would coincide. This does not mean, however, that in those situations, $\beta(H_0, H_i, \alpha)$ of the Type II test would be the β of the Type I test. Given H_0 and the H_i's, it would be possible to find a β for Type I and a μ for Type II such that the same critical region for H_0 against the whole class $H_i \varepsilon \Omega$ could be looked upon as $U_{H_i}\omega(H_0, H_i, \beta)$ in relation to the first type and also as $U_{H_i}\omega^*(H_0, H_i, \mu)$ in relation to the second type.

The following theoretical question or group of questions now under investigation is extremely important. Under what *general* restrictions on the probability law of $\underline{x}$ and on H_0 and $H_i \varepsilon \Omega$ would either or both of the tests be nontrivial (in the sense discussed in Section 3) and usable (in the sense of having a distribution problem amenable to tabulation), and unbiased (against all relevant alternatives) and/or admissible and/or reasonably powerful (in the sense of having not too bad a power against all relevant alternatives)? So far as the author is aware, these questions have not yet been adequately discussed in a *general* manner (let alone been answered) even for the likelihood ratio or Type II test (which has so long been extensively used in practice), and no attempt will be made in this paper to discuss these questions. The advantage, however, of having two such heuristic principles (with the possibility of having two different tests in many situations) is that it gives us more elbow room than we would have had with one such principle, in the matter of construction of nontrivial, usable and "pretty good" tests.

4. Extended Type I test (and an obvious property of it). Consider a composite hypothesis H_0 against a set of composite alternatives $H_i \varepsilon \Omega$, $(i = 1, 2, \cdots)$. It often happens, as for example in the three broad situations discussed in Section 5, that, while there are similar regions for H_0, there is among these no most powerful (bisimilar) region for H_0 against any $H_i (i = 1, 2, \cdots)$, but that we have instead the following situation. Suppose we have composite hypotheses $H_{0j} (j = 1, 2, \cdots)$ such that $\bigcap_j H_{0j} \equiv H_0$ and composite alternatives $H_{ij} (i = 1, 2, \cdots; j = 1, 2, \cdots)$ such that $\bigcap_j H_{ij} \equiv H_i$. Notice that H_{0j} and H_{ij} have more unspecified elements than H_0 and H_i respectively. It may well be that we have (as in the cases discussed in Section 5) not only similar regions for H_{0j} but also, among these, a most powerful (bisimilar) region for H_{0j} against any H_{ij} (one for each i with $j = 1, 2, \cdots$; and $i = 1, 2, \cdots$). Consider critical regions $\omega(H_{0j}, H_{ij}, \beta)$ of size β each. Then by our test procedure, over $\bigcap_j \bigcap_i$ of $\bar{\omega}(H_{0j}, H_{ij}, \beta)$ (which we call $\bigcap_{ji}$ for simplicity), we are anyway accepting $\bigcap_j H_{0j}$, that is, H_0 and over its complement $U_j U_i \omega(H_{0j}, H_{ij}, \beta)$ we are rejecting at least one H_{0j} and therefore H_0 itself. Suppose we set this up as a heuristic test for H_0 against the whole class $H_i \varepsilon \Omega$. Then the critical region will be $U_j U_i \omega(H_{0j}, H_{ij}, \beta)$ or U_{ji} of size α, given by

$$P(\underline{x} \varepsilon U_{ji} \mid H_0) = \alpha \tag{4.1}$$

so that $\alpha \equiv \alpha(H_0, \beta)$ and $\beta \equiv \beta(H_0, \alpha)$. As before, nontriviality will be assumed, and it is easy to check that we shall have for all i and j the following inequality:

$$\beta < P(H_{0j}, H_{ij}, \beta) \leqq P(U_{ji}, H_i, \alpha) \leqq 1. \tag{4.2}$$

It may be noted that while $\omega(H_{0j}, H_{ij}, \beta)$, a bisimilar region of size β for H_{0j} against H_{ij}, is independent of the unspecified elements of H_{0j} and H_{ij} and while the location of U_{ji} must be and its *size might* be (as indeed it is for all the cases considered in Section 5) independent of the unspecified elements of H_{0j} and H_{ij}, $P(H_{0j}, H_{ij}, \beta)$, but *might* involve the unspecified elements of H_{ij} and $P(H_0, H_i, \alpha)$ involve those of H_i. As observed in Section 3, the lower bound to the power of the test, given by (4.2), while it is in general easily available, is, at the same time, much too crude. With more knowledge of the probability law a much closer lower bound can often be found as will be exemplified in later sections.

5. Application to three multivariate problems.

5.1. *Statement of the problems.* Three different types of hypotheses will be discussed here, namely, (i) the hypothesis of equality of covariance matrices of two p-variate normal populations, (ii) the hypothesis of equality of k means for each of p variates for k p-variate normal populations with the same covariance matrix (which is *formally* tied up with the general problem of testing a linear hypothesis), and (iii) the hypothesis that in a $(p_1 + p_2)$-variate normal population the set of, say, the first p_1 variates is uncorrelated with the set of the last p_2 variates. In symbols, using the notation given in Section 2, we can rewrite these hypotheses as (i) $H(\Sigma_1 = \Sigma_2)(\equiv H_0)$ against *all* $H(\Sigma_1 \neq \Sigma_2)(\equiv H)$, (ii) $H(\xi_1 = \xi_2 \cdots = \xi_k)(\equiv H_0)$ (assuming a common Σ) against *all* $H(\neq H_0)$ (assuming again a common Σ) and (iii) $H(\Sigma_{12} = 0)(\equiv H_0)$ against *all* $H(\Sigma_{12} \neq 0)$ $(\equiv H)$, where the $(p_1 + p_2)$ variates have a covariance matrix Σ of the following structure:

$$\Sigma = \begin{array}{cc} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} & \begin{array}{l} p_1 \\ p_2 \end{array} \\ \begin{array}{cc} p_1 & \quad p_2 \end{array} & \end{array} . \tag{5.1}$$

5.2. *Direct Type I construction not possible.* It is well known that there are infinitely many similar regions for each of the above composite hypotheses but no most powerful (bisimilar) region for $H(\Sigma_1 = \Sigma_2)$ against any *specific* $H(\Sigma_1 \neq \Sigma_2)$ or for $H(\xi_1 = \cdots = \xi_k)(\equiv H_0)$ against any *specific* H violating H_0 or for $H(\Sigma_{12} = 0)$ against any *specific* $H(\Sigma_{12} \neq 0)$, so that direct Type I construction will not work here.

5.3. *Reduction to pseudo-univariate and pseudo-bivariate problems.* At this point suppose that, starting from an $\underset{\sim}{x}(: p \times 1)$ which is $N(\underset{\sim}{\xi}, \Sigma)$, we consider a linear compound of $\underset{\sim}{x}$, namely $\underset{\sim}{\mu}'\underset{\sim}{x}$ (with an arbitrary constant, that is, nonstochastic $\underset{\sim}{\mu}'(:1 \times p)$ of nonzero modulus) which is a scalar well known to be $N(\underset{\sim}{\mu}'\xi, \underset{\sim}{\mu}'\Sigma\underset{\sim}{\mu})$. Note that $\underset{\sim}{\mu}'\underset{\sim}{\xi}$ and $\underset{\sim}{\mu}' \Sigma\underset{\sim}{\mu}$ are also scalars. Suppose further that we also start from

$$\underset{\sim}{x} = \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} \begin{array}{l} p_1 \\ p_2 \end{array} = N\left(\begin{pmatrix} \underset{\sim}{\xi}_1 \\ \underset{\sim}{\xi}_2 \end{pmatrix} : \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right),$$

and consider linear compounds $\underset{\sim}{\mu}_1'\underset{\sim}{x}_1$ and $\underset{\sim}{\mu}_2'\underset{\sim}{x}_2$ (where $\underset{\sim}{\mu}_1(:p_1 \times 1)$ and $\underset{\sim}{\mu}_2(:p_2 \times 1)$ are each nonnull); then these two scalars are well known to be distributed as a bivariate normal with a correlation coefficient

$$\rho(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2) \equiv \rho_{12} = \underset{\sim}{\mu}_1'\Sigma_{12}\underset{\sim}{\mu}_2/[(\underset{\sim}{\mu}_1'\Sigma_{11}\underset{\sim}{\mu}_1)^{\frac{1}{2}}(\underset{\sim}{\mu}_2'\Sigma_{22}\underset{\sim}{\mu}_2)^{\frac{1}{2}}]. \tag{5.3.1}$$

Now suppose that, in place of (i), (ii) and (iii) of 5.1, we consider respectively

(iv) $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})(\equiv H_{0\underset{\sim}{\mu}})$ against *all* $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} \neq \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})(\equiv H_{\underset{\sim}{\mu}})$, ($\underset{\sim}{\mu}$ fixed),
(v) $H(\underset{\sim}{\mu}'\xi_1 = \cdots = \underset{\sim}{\mu}'\xi_k)(\equiv H_{0\underset{\sim}{\mu}})$ against *all* $H_{\underset{\sim}{\mu}}(\neq H_{0\underset{\sim}{\mu}})$, ($\underset{\sim}{\mu}$ fixed) and
(vi) $H(\underset{\sim}{\mu}_1'\Sigma_{12}\underset{\sim}{\mu}_2 = 0)(\equiv H_{0\underset{\sim}{\mu}_1\underset{\sim}{\mu}_2})$ against *all* $H(\underset{\sim}{\mu}_1'\Sigma_{12}\underset{\sim}{\mu}_2 \neq 0)(\equiv H_{\underset{\sim}{\mu}_1\underset{\sim}{\mu}_2})(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2$ fixed).

We now consider the *totality of all* nonnull $\underset{\sim}{\mu}$ for (iv) and (v) and *all* nonnull $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ for (vi). Notice that (i) $\bigcap_{\underset{\sim}{\mu}} H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu}) = H(\Sigma_1 = \Sigma_2)$, (ii) $\bigcap_{\underset{\sim}{\mu}} H(\underset{\sim}{\mu}'\xi_1 = \cdots = \underset{\sim}{\mu}'\underset{\sim}{\xi}_k) = H(\xi_1 = \cdots = \underset{\sim}{\xi}_k)$ and (iii) $\bigcap_{\underset{\sim}{\mu}_1\underset{\sim}{\mu}_2} H(\underset{\sim}{\mu}_1'\Sigma_{12}\underset{\sim}{\mu}_2 = 0) = H(\Sigma_{12} = 0)$. We could have worked in terms of any subset of such $\underset{\sim}{\mu}$'s which leads by intersection to the same H_0, but this we do not do here. It may be noted that by the procedure to be used here, apart from measure-theoretic difficulties which, however, do not arise in these applications, the total set of

$\underset{\sim}{\mu}$'s or any subset of it (of the kind considered) will *uniquely* define an extended Type I test *associated with the total set or with that particular subset.* Next suppose that, in the alternative, under (iv), (v) and (vi), we substitute "*specific*" for "*all*" and thus have three new situations (vii), (viii) and (ix). It is well known that for each of the situations (vii), (viii) and (ix) we have *one most powerful (bisimilar) region,* so that from these we can construct respective *Type I regions* for the univariate situations (iv) and (v) and the bivariate situation (vi), and from these *Type I tests* we can try to construct the respective *extended Type I tests* for the situations (i), (ii) and (iii). This ties up (in the Section 4) the two p-variate problems (i) and (ii) with the two univariate problems (iv) and (v), and the $(p_1 + p_2)$-variate problem with the bivariate problem (vi).

5.4. *A useful notation and reduction.* For an observation matrix $X(:p \times n)$ with elements $x_{i\lambda}(i = 1, 2, \cdots, p; \lambda = 1, 2, \cdots, n)$, les us put $\bar{x}_i \equiv \sum_{\lambda=1}^{n} x_{i\lambda}/n$, $(i = 2, \cdots, p)$ and $\underset{\sim}{x}'(\equiv \bar{x}_1, \cdots, \bar{x}_p)$. Then the covariance matrix $S(:p \times p)$ will be given by $(n - 1)S = XX' - n\underset{\sim}{x}\underset{\sim}{x}'$. Now suppose that in the situation (i) we have two observation matrices $X_r(:p \times n_r)$, $p \leqq n_r - 1$, two mean vectors $\underset{\sim}{x}_r:(p \times 1)$ and two covariance matrices $S_r(:p \times p)$ such that $(n_r - 1)\, S_r = X_r X_r' - n_r \underset{\sim}{x}_r \underset{\sim}{x}_r'$ $(r = 1, 2)$, so that S_r is always at least p.s.d. In situation (ii) assume that we have k observation matrices $X_r(:p \times n_r)$; mean vectors $\underset{\sim}{x}_r(:p \times 1)$; a grand mean vector $\underset{\sim}{x}(:p \times 1)$ such that $\underset{\sim}{x} = \Sigma n_r \underset{\sim}{x}_r/n$, where $n \equiv \sum_{r=1}^{k} n_r$ and $p \leqq n - k$; a covariance matrix of means $S^*(:p \times p)$ defined by $(k - 1)S^* = XX' - n\underset{\sim}{x}\,\underset{\sim}{x}'$, where $X(:p \times k) \equiv (\sqrt{n_1}\underset{\sim}{x}_1 \cdots \sqrt{n_k}\,\underset{\sim}{x}_k)$; k covariance matrices $S_r(:p \times p)$ such that $(n_r - 1)S_r = X_r X_r' - n_r \underset{\sim}{x}_r \underset{\sim}{x}_r'$; a pooled covariance matrix $S(:p \times p)$ such that $(n - k)S = \sum_{r=1}^{k}(n_r - 1)S_r$ so that both S^* and S must be always at least p.s.d. Finally in situation (iii) suppose that we have an observation matrix X and a mean vector $\underset{\sim}{x}$ given by:

$$X \equiv \underset{n}{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}} \begin{matrix} p_1 \\ p_2 \end{matrix} \quad \text{and} \quad \underset{\sim}{x} \equiv \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix},$$

and a covariance matrix $S(:(p_1 + p_2) \times (p_1 + p_2))$ given by:

$$(n - 1)S \equiv (n - 1) \begin{array}{c} \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} \\ \begin{matrix} p_1 & \quad p_2 \end{matrix} \end{array} \equiv \begin{pmatrix} X_1 X_1' & X_1 X_2' \\ X_2 X_1' & X_2 X_2' \end{pmatrix} - n \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} (\underset{\sim}{x}_1' \underset{\sim}{x}_2').$$

Here we observe that S must be always at least p.s.d. and also assume that $p_1 \leqq p_2$ and $p_1 + p_2 \leqq n - 1$.

5.5 *Type I tests for the situations (iv), (v) and (vi).*

(iv) Put $F_{\underset{\sim}{\mu}} = \underset{\sim}{\mu}' S_1 \underset{\sim}{\mu} / \underset{\sim}{\mu}' S_2 \underset{\sim}{\mu}$ and notice that, at a level β, for $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})$ $(\equiv H_{0\underset{\sim}{\mu}})$ against all $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} > \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})$ we have the one-sided uniformly most powerful (bisimilar) region: $F_{\underset{\sim}{\mu}} \geqq F_0$, and for $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})(\equiv H_{0\underset{\sim}{\mu}})$ against all $H(\underset{\sim}{\mu}'\Sigma_1\underset{\sim}{\mu} < \underset{\sim}{\mu}'\Sigma_2\underset{\sim}{\mu})$ we have the one-sided uniformly most powerful region:

$F_\mu \leqq F_0'$, where F_0 and F_0' are given by: $P(F_\mu \geqq F_0 \mid H_{0\mu}) = P(F_\mu \leqq F_0' \mid H_{0\mu}) = \beta$. Notice that this F_μ has the ordinary F-distribution with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom. The Type I critical region will now be of size 2β, being given by

$$\omega_\mu: (F_\mu \geqq F_0) U (F_\mu \leqq F_0'), \quad \text{or} \quad : F_\mu \geqq F_0 \quad \text{or} \quad \leqq F_0'. \tag{5.5.1}$$

For $n_1 = n_2$ this will be an unbiased critical region, but, for $n_1 \neq n_2$, this will be biased for certain small deviations and unbiased for all large deviations from the hypothesis. In any case, in this situation it is possible to construct a better (but slightly more difficult test) which will not be discussed here.

(v) For $H(\mu'\xi_1 = \cdots = \mu'\xi_k)(\equiv H_{0\mu})$ against any *specific* $H_\mu(\neq H_{0\mu})$ there is the most powerful (bisimilar) critical region (discussed in Section 3) (of size, say, γ) which is a one-sided t-region, and by taking the union of these regions (for fixed μ but by variations over $\xi_1, \cdots, \xi_k$), we have the Type I region given by

$$F_\mu \equiv \mu' S^* \mu / \mu' S \mu \geqq F_0, \tag{5.5.2}$$

when F_0 is obtained from $P(F_\mu \geqq F_0 \mid H_{0\mu}) = \beta$.
This is also well known to be a Type II or likelihood ratio test having in this situation various good properties (including unbiasedness and admissibility). Notice that this F_μ has the ordinary F-distribution with $(k - 1)$ and $(n - k)$ degrees of freedom.

(vi) Put $r_{\mu_1\mu_2} \equiv \mu_1' S_{12} \mu_2 / (\mu_1' S_{11} \mu_1)^{\frac{1}{2}} (\mu_2' S_{22} \mu_2)^{\frac{1}{2}}$ and notice that, at a level β, for $H(\mu_1'\Sigma_{12}\mu_2 = 0)$ $(\equiv H_{0\mu_1\mu_2})$ against all $H(\mu_1'\Sigma_{12}\mu_2 > 0)$ we have the one-sided uniformly most powerful (bisimilar) region: $r_{\mu_1\mu_2} \geqq r_0$ and for $H_{0\mu_1\mu_2}$ against all $H(\mu_1'\Sigma_{12}\mu_2 < 0)$ we have the one-sided uniformly most powerful (bisimilar) region $r_{\mu_1\mu_2} \leqq -r_0$, where r_0 is given by:

$$P(r_{\mu_1\mu_2} \geqq r_0 \mid H_{0\mu_1\mu_2}) = \beta. \tag{5.5.3}$$

Notice that $r_{\mu_1\mu_2}$ has the distribution of the ordinary total correlation coefficient on a sample of size n. The Type I critical region will be of size 2β, being given by

$$\omega_{\mu_1\mu_2}: (r_{\mu_1\mu_2} \geqq r_0)\ U (r_{\mu_1\mu_2} \leqq -r_0), \tag{5.5.4}$$

that is, $|r| \geqq r_0 |$ or $r^2 \geqq r_0^2$.
This is well known to be also a Type II or likelihood ratio region having in this situation various good properties (including unbiasedness and admissibility).

5.6. *Actual construction of extended Type I tests for the situations (i), (ii) and (iii).*

(i) By the test procedure (5.5.1), over $F_0' < F_\mu < F_0$ we accept $H(\mu'\Sigma_1\mu = \mu'\Sigma_2\mu)$ so that over $\bigcap_\mu[F_0' < F_\mu = \mu' S_1 \mu / \mu' S_2 \mu < F_0]$ we accept $\bigcap_\mu H(\mu'\Sigma_1\mu = \mu'\Sigma_2\mu) \equiv H(\Sigma_1 = \Sigma_2) \equiv H_0$, and thus over its complement $U_\mu[F_\mu \geqq F_0 \text{ or } \leqq F_0']$ we reject H_0. This may thus be set up as the extended Type I test. To obtain $U_\mu [F_\mu \geqq F_0 \text{ or } \leqq F_0']$ we note that a particular set of observations, that is, a particular set of (S_1, S_2) would belong to the union if for that (S_1, S_2) there were at least one μ such that $F_\mu \geqq F_0$ or $\leqq F_0'$. It is thus easy to check that $U_\mu[F_\mu \geqq$

F_0 or $\leqq F_0']$ is precisely equivalent to: the largest $F_{\mu} \geqq F_0$ and/or the smallest $F_{\mu} \leqq F_0'$, the "largest" and the "smallest" being under variation of μ (for a given set of S_1, S_2). Now, given (S_1, S_2), the largest and smallest value of $\mu' S_1 \mu / \mu' S_2 \mu$ are easily seen to be the largest and smallest roots, say θ_p and θ_1, of the p-th degree determinantal equation in θ

$$(5.6.1) \qquad | S_1 - \theta S_2 | = 0,$$

all the p roots $\theta_1, \theta_2, \cdots, \theta_p$ being in this situation a.e. positive, since S_1 and S_2 are by the definitions and assumptions of subsection 5.4 of Section 5, a.e., p.d. (each of rank p). Starting out from the Type I test (5.5.1) for $H_{0\mu}$ we have for $H(\Sigma_1 = \Sigma_2)$ the extended Type I critical region

$$(5.6.2) \qquad \theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0'.$$

To determine the size of this critical region, or more properly, given the size α, to find F_0 and F_0', we have to have the joint distribution of $(\theta_1, \theta_2, \cdots, \theta_p)$ on the null hypothesis $H(\Sigma_1 = \Sigma_2)$ which was obtained in 1939 by a number of workers [3], [4], [7], [10] and which was found to be independent of the common value of $\Sigma_1 = \Sigma_2$ and also of ξ_1 and ξ_2, that is, of all nuisance parameters. Starting from the joint distribution of $(\theta_1, \cdots, \theta_p)$ on the null hypothesis, we can obtain, by a technique given in earlier papers [9], [12], the joint distribution of (θ_1, θ_p), from which F_0 and F_0' will be available, in terms of α, by using

$$(5.6.3) \qquad P(\theta_p \geqq F_0 \mid \Sigma_1 = \Sigma_2) = P(\theta_1 \leqq F_0' \mid \Sigma_1 = \Sigma_2), \text{ and } P(\theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0' \mid \Sigma_1 = \Sigma_2) = \alpha.$$

(ii) By the test procedure (5.5.2), over $F_{\mu} \equiv \mu' S^* \mu / \mu' S \mu < F_0$ accept $H(\mu'\xi_1 = \cdots = \mu'\xi_k)$, so that over $\bigcap_{\mu}[F_{\mu} < F_0]$ we accept $\bigcap_{\mu} H(\mu'\xi_1 = \cdots \mu'\xi_k) \equiv H(\xi_1 = \cdots = \xi_k) \equiv H_0$, and over its complement $U_{\mu}[F_{\mu} \geqq F_0]$ we reject H_0. We set it up as the extended Type I test for H_0, and note, as before, that $U_{\mu}[F_{\mu} \equiv \mu' S^* \mu / \mu' S \mu \geqq F_0]$ is precisely equivalent to: the largest $\mu' S^* \mu / \mu' S \mu \geqq F_0$, the "largest" being under variation of μ (for a given set of observations, that is, for a given set of S^* and S). As before, given S^* and S, the largest value of $\mu' S^* \mu / \mu' S \mu$ is checked to be the largest root θ_q of the p-th degree determinantal equation in θ

$$(5.6.4) \qquad | S^* - \theta S | = 0.$$

From the definitions and assumptions of (subsection 5.4) of Section 5, it is easy to check that S is, a.e., p.d. while S^* is, a.e., at least p.s.d. of rank $q \equiv \min(p, k - 1)$. It will of course be, a.e., p.d. if $p \leqq k - 1$. In any case we can say that, of the p roots of (5.41), $p - q$ will be always zero, while q roots, to be called $\theta_1, \cdots, \theta_q$, will be, a.e., positive, where $q \equiv \min(p, k - 1)$, so that $0 < \theta_1 \leqq \cdots \leqq \theta_q < \infty$ (suppose). The extended Type I critical region for H_0 is thus

$$(5.6.5) \qquad \theta_q \geqq F_0 .$$

To determine the size of the region, or rather, given the size α of (5.6.5), to determine F_0, we observe what was noted in the earlier papers [4], [7], [10], namely, that the joint distribution of $(\theta_1, \cdots, \theta_q)$ (on the null hypothesis) in this case is exactly of the same form as that of $(\theta_1, \cdots, \theta_p)$ of the previous case (on the null hypothesis in that situation) and that therefore the distribution of any root, say the largest, will come through by the same technique as was mentioned for the previous case and will also be independent of all nuisance parameters. We shall thus have F_0 given, in terms of α, by,

$$P(\theta_q \geqq F_0 \mid \underline{\xi}_1 = \cdots = \underline{\xi}_k) = \alpha. \tag{5.6.6}$$

For $k = 2$ we shall have $q = 1$, so that there will be just one nontrivial sample root $\theta_q(\equiv\theta$ suppose), and just one nontrivial population root $\Theta_r(\equiv\Theta$ suppose) (which will be zero on the null hypothesis and $\neq 0$, on the nonnull hypothesis). This θ is easily checked to be Hotelling's T^2 and its distribution both on the null and nonnull hypothesis are well known [1], [5] and relatively easy, so that (5.6.5) and (5.6.6) happen to be computationally much simpler in this situation.

(iii) By the test procedure (5.5.4), over $r^2_{\mu_1\mu_2} \equiv (\mu_1'S_{12}\mu_2)^2/[(\mu_1'S_{11}\mu_1)(\mu_2'S_{22}\mu_2)] \geqq r_0^2$, we reject $H(\mu_1'S_{12}\mu_2 = 0) \equiv H_{0\mu_1\mu_2}$ and over its complement accept this hypothesis, so that over $\bigcap_{\mu_1\mu_2}[r^2_{\mu_1\mu_2} < r_0^2]$ we accept $\bigcap_{\mu_1\mu_2}H(\mu_1'\Sigma_{12}\mu_2 = 0) \equiv H_0$, and over its complement $U_{\mu_1\mu_2}[r^2_{\mu_1\mu_2} \geqq r_0^2]$ reject H_0. We set this up as the extended Type I test for H_0 and note that

$$U_{\mu_1\mu_2}[r^2_{\mu_1\mu_2} \equiv (\mu_1'S_{12}\mu_2)^2/(\mu_1'S_{11}\mu_1)(\mu_2'S_{22}\mu_2) \geqq r_0^2]$$

is exactly equivalent to: the largest value of

$$(\mu_1'S_{12}\mu_2)^2/(\mu_1'S_{11}\mu_1)(\mu_2'S_{22}\mu_2) \geqq r_0^2,$$

the "largest" being under variation of μ_1 and μ_2 (for a given set of observations, that is, for a given set of S_{11}, S_{22} and S_{12}). As before, the largest value of this expression is checked to be the largest root θ_{p_1} of the p_1st degree determinantal equation in θ

$$| \theta S_{11} - S_{12}S_{22}^{-1}S_{12}' | = 0. \tag{5.6.7}$$

From the definitions and assumptions of subsection 5.4 of Section 5, it is easy to see that S and, therefore, S_{11} and S_{22} are, a.e., p.d. and S_{12} is, a.e., of rank p_1. Under these conditions it is well known and proved in a number of places [6], [15] that the p_1 roots of (5.6.7) will all, a.e., lie between 0 and 1, satisfying, say, $0 < \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_{p_1} < 1$. The extended Type I region for H_0 is thus

$$\theta_{p_1} \geqq r_0^2. \tag{5.6.8}$$

To determine the size of this region, or rather, given the size α, to determine r_0^2, we observe that the joint distribution of $(\theta_1, \cdots, \theta_{p_1})$ on the null hypothesis in this case goes over (under a simple transformation from cosine to cotangent) into that of the joint distribution of the roots (on the respective null hypotheses) in the two previous cases and the same technique for finding the distribution of the largest root also goes through. As before, this distribution will also be independent of all nuisance parameters. We shall thus have r_0^2 given, in terms of α, by,

$$P(\theta_{p_1} \geqq r_0^2 \mid \Sigma_{12} = 0) = \alpha. \tag{5.6.9}$$

6. Lower bounds of the powers of the test regions (5.6.2), (5.6.5) and (5.6.8) for the hypotheses (i), (ii), and (iii).

6.1. *Observations on the actual power functions.*

(i) It is well known that on the nonnull hypothesis the joint distribution of $(\theta_1, \cdots, \theta_p)$ of (5.6.1) (and hence of $(\theta_1$ and $\theta_p))$ also involves as parameters only the p roots $\Theta_1, \cdots, \Theta_p$ of the population determinantal equations in Θ,

$$| \Sigma_1 - \Theta\Sigma_2 | = 0. \tag{6.1.1}$$

(Notice that, assuming Σ_1 and Σ_2 to be both p.d., these roots will all be positive and they will all be unity if and only if $\Sigma_1 = \Sigma_2$, that is, on the null hypothesis in this situation.) The exact distribution of $(\theta_1, \cdots, \theta_p)$ or of (θ_1, θ_p) on the nonnull hypothesis will be quite complicated and whatever reduction is already known to be possible [11], will not be discussed here. We shall merely write the power function formally as:

$$\begin{aligned} &P[\theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0' \mid \Sigma_1 \neq \Sigma_2] \\ &\qquad \equiv P\{\alpha; n_1, n_2, p; \Theta_1, \Theta_2, \cdots, \Theta_p\}, \end{aligned} \tag{6.1.2}$$

to indicate on which parameters the power depends.

(ii) To discuss the power function of the region (5.6.5), we use the convenient notation: $\underset{\sim}{\xi} \equiv \sum_{r=1}^{k} n_r\underset{\sim}{\xi}_r/n$; $\xi(:p \times k) \equiv (\sqrt{n_1}\underset{\sim}{\xi}_1, \cdots, \sqrt{n_k}\underset{\sim}{\xi}_k)$; $(k - 1)\Sigma^* \equiv \xi\xi' - n\underset{\sim}{\xi}\underset{\sim}{\xi}'$; denote by Σ the (assumed) common p.d. covariance matrix of the k populations. We note that $\Sigma^*(:p \times p)$ is p.s.d. (and might also be p.d.) of rank $r \leqq \min(p, k - 1)$, where r is the rank of the matrix,

$$(\sqrt{n_1}(\underset{\sim}{\xi}_1 - \underset{\sim}{\xi}), \cdots, \sqrt{n_k}(\underset{\sim}{\xi}_k - \underset{\sim}{\xi})).$$

Notice that the rank of this matrix must be $\leqq \min(p, k - 1)$. Notice further that Σ^* will be zero if and only if $\underset{\sim}{\xi}_1 = \underset{\sim}{\xi}_2 = \cdots = \underset{\sim}{\xi}_k$, that is, on the null hypothesis in this situation. We next observe, as is well known, that on the nonnull hypothesis the joint distribution of $(\theta_1, \cdots, \theta_q)$ of (5.6.4) (and also of θ_q) will involve as parameters only the $r(\leqq q \equiv \min(p, k - 1))$ positive (the $p - r$ others being zero) roots of the p-th degree population determinantal equation in Θ,

$$| \Sigma^* - \Theta\Sigma | = 0. \tag{6.1.3}$$

As in the previous cases, so also here, the exact distribution of $(\theta_1, \cdots, \theta_q)$ or of θ_q on the nonnull hypothesis will be quite complicated *and also* different from that of the previous situation and whatever reduction is already known to be possible will, as before. not be discussed here. We shall again formally write the power function as

$$P[\theta_q \geqq F_0 \mid \text{under violation of } H(\xi_1 = \cdots = \xi_k)] \equiv P\{\alpha; n, k, p; \Theta_1, \cdots, \Theta_r\}, \tag{6.1.4}$$

to indicate the dependence on the relevant parameters. When $k = 2$ we have $q = 1$, $r = 1$, and in this case (6.1.4) will be the power function of Hotelling T-test, which is computationally quite manageable.

(iii) To discuss the power function of the region (5.6.8), we observe what is well known, namely that, on the nonnull hypothesis, the joint distribution of $(\theta_1, \cdots, \theta_{p_1})$ of (5.6.7) involves as parameters only the roots of the p_1-th degree population determinantal equation in Θ,

$$| \Theta\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' | = 0. \tag{6.1.5}$$

Assuming a p.d. Σ, it is also known that Σ_{11} and Σ_{22} are both p.d., all roots being less than 1 and q roots being positive and $p_1 - q$ being zero, where q is the rank of $\Sigma_{12}(q \leqq p_1 \leqq p_2)$. We write them as: $0 < \Theta_1 \leqq \cdots \leqq \Theta_q < 1$. We shall not further discuss the complicated nonnull distribution of $(\theta_1, \cdots, \theta_{p_1})$ or of θ_{p_1}, but merely write down formally the power function of the critical region (5.6.8) as,

$$P[\theta_{p1} \geqq r_0^2 \mid \Sigma_{12} \neq 0]. \equiv P\{\alpha; n, p_1, p_2; \Theta_1, \cdots, \Theta_q\}, \tag{6.1.6}$$

to indicate the dependence on the relevant parameters.

Although the exact nonnull distributions and hence the exact power functions would be quite complicated in all the foregoing cases we could, if we wanted to, obtain lower bounds, by using (4.2) and noting that the nonnull distribution for the univariate situations (iv) and (v) associated with (i) and (ii), and the bivariate situation (vi) associated with (iii), are all known in computationally manageable forms. But it is possible, as is shown in the next two subsections (6.2) and (6.3), to obtain much closer lower bounds to the power functions (6.1.2), (6.6.4) and (6.1.6). This is accomplished as follows.

6.2. *On invariance and independence.* It is well known [15] that

(i) the roots of (5.6.1) are invariant under the transformation

$$S_1(:p \times p) = \mu(:p \times p)V_1(:p \times p)\mu'(:p \times p) \quad \text{and}$$
$$S_2(:p \times p) = \mu(:p \times p)V_2(:p \times p)\mu'(:p \times p),$$

when μ is any constant (i.e., nonstochastic) nonsingular transformation matrix,

(ii) the roots of (5.6.4) are invariant under the transformation:

$$S^*(:p \times p) = \mu(:p \times p)V^*(:p \times p) \times \mu'(:p \times p) \text{ and}$$
$$S(:p \times p) = \mu(:p \times p)V(:p \times p)\mu'(:p \times p),$$

where μ is any constant nonsingular transformation matrix, and finally

(iii) the roots of (5.6.7) are invariant under the transformation:

$$S_{11}(:p_1 \times p_1) = \mu_1(:p_1 \times p_1)V_{11}(:p_1 \times p_1) \times \mu_1'(:p_1 \times p_1),$$
$$S_{22}(:p_2 \times p_2) = \mu_2(:p_2 \times p_2)V_{22}(:p_2 \times p_2) \times \mu_2'(:p_2 \times p_2)$$

and

$$S_{12}(:p_1 \times p_2) = \mu_1(:p_1 \times p_1)V_{12}(:p_1 \times p_2) \times \mu_2'(:p_2 \times p_2),$$

where μ_1 and μ_2 are any two constant nonsingular transformation matrices.

We next notice that

(i) there exists a nonsingular transformation matrix (not necessarily unique),

$$\mu(:p \times p) \equiv (\mu_1, \cdots, \mu_p)p,$$

under which $\mu\Sigma_1\mu' = D_\Theta$ and $\mu\Sigma_2\mu' = I(p)$ (where D_Θ is a $p \times p$ diagonal matrix whose diagonal elements are $\Theta_1, \cdots, \Theta_p$) and which transforms the p original variates into p new variates distributed in a canonical form, so that, for this set of p μ_i's$(i = 1, 2, \cdots, p)$, $(\mu_i'S_1\mu_i/\mu_i'S_2\mu_i)/(\mu_i'\Sigma_1\mu_i/\mu_i'\Sigma_2\mu_i)$, that is, $(\mu_i'S_1\mu_i/\mu_i'S_2\mu_i)\ \Theta_i(i = 1, \cdots, p)$ will be distributed as p independent F's, each with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom,

(ii) there exists a nonsingular matrix (not necessarily unique), $\mu(:p \times p) \equiv (\mu_1, \cdots, \mu_p)p$, under which $\mu\Sigma^*\mu' = D_\Theta$ and $\mu\Sigma\mu' = I(p)$ (where D_Θ is a diagonal matrix, of whose p diagonal elements, $p - r$ are exactly zero, while the rest, r in number, are $\Theta_1, \cdots, \Theta_r > 0$), and, furthermore, that this transforms the p original variates into p new variates distributed in a canonical form, so that for this set of p μ_i's$(i = 1, 2, \cdots, p)$, $(\mu_i'S^*\mu_i/\mu_i'S\mu_i)(i = 1, 2, \cdots, p)$ will be distributed as p independent F's each with $(k - 1)$ and $(n - k)$ degrees of freedom. We note that out of these p F's, $p - r$ are necessarily central F's (i.e., with "deviation parameters" equal to zero) and r F's are noncentral with "deviations parameters", $(\Theta_1, \cdots, \Theta_r)$ and

(iii) there exist nonsingular matrices (none necessarily unique),

$$\mu_1(:p_1 \times p_1) \equiv (\mu_{11} \cdots \mu_{p_1 1})p_1,$$

$$\mu_2(:p_2 \times p_2) \equiv (\mu_{12} \cdots \mu_{p_2 2})p_2,$$

under which $\mu_1\Sigma_{11}\mu_1' = I(p_1)$, $\mu_2\Sigma_{22}\mu_2' = I(p_2)$ and

$$\mu_1\Sigma_{12}\mu_2' = (D_{\sqrt{\Theta}}\ 0)p_1$$

(where $D_{\sqrt{\Theta}}$ is a $p_1 \times p_1$ diagonal matrix of whose diagonal elements, $p_1 - q$ are zero and the rest are nonzero, being $\Theta_1, \cdots, \Theta_q$), and which transforms the original $(p_1 + p_2)$ variates into two new sets of p_1 and p_2 variates, jointly distributed in a canonical form with covariance matrix:

$$\begin{array}{cc} \begin{pmatrix} I(p_1) & (D_{\sqrt{\Theta}}\ 0) \\ \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} & I(p_2) \end{pmatrix} & \begin{array}{l} p_1 \\ p_1 \\ p_2 - p_1 \end{array} \\ \begin{array}{ccc} p_1 & p_1 & p_2 - p_1. \end{array} & \end{array}$$

This means that from the sets $\mu_{i1}(i = 1, 2, \cdots, p_1)$ and $\mu_{j2}(j = 1, 2, \cdots, p_2)$ it is possible to pick out linked μ_{i1} and $\mu_{i2}(i = 1, 2, \cdots, p_1)$ such that

$(\mu_{i1}' S_{12} \mu_{i2})^2 / (\mu_{i1}' S_{11} \mu_{i1})(\mu_{i2}' S_{22} \mu_{i2})(i = 1, 2, \cdots, p_1)$ are distributed as the squares of p_1 independent correlation coefficients r_i with $(n - 2)$ degrees of freedom each, the distributions involving $\Theta_i \equiv \rho_i^2 (i = 1, 2, \cdots, q \leqq p_1)$ as "deviation parameters". The absolute value of the total correlation coefficient will be indicated by enclosing the correlation in vertical bars. It is the distribution of this, that is, the distribution of multiple correlation when $p = 2$, that will come into the picture. It is possible to go even beyond this and pick out linked μ_{i1} and $\mu_{i2}(i = 1, 2, \cdots, p_1 - 1)$, and at the last stage a $\mu_{p_1 1}$ linked with a set of $(p_2 - p_1 + 1)$ μ_{i2}'s$(i = p_1, p_1 + 1, \cdots, p_2)$, such that there are p_1 independently distributed | correlations |, of which $(p_1 - 1)$ are | total correlations |, and the last one is a multiple correlation between the p_1th variate of the first p_1-set and the $(p_1, p_1 + 1, \cdots, p_2)$ variates of the second p_2-set. The deviation parameters being $\Theta_i (0 < \Theta_1 \leqq \cdots \leqq \Theta_q < 1)$, we could so arrange that the first $p_1 - q$ sample (total) | correlations | had zero deviation parameters to go with, the next $q - 1$ sample (total) | correlations | had respective (and one each) deviation parameters $(\Theta_1, \cdots, \Theta_{q-1})$ to go with and the last sample (multiple) correlation had Θ_q to go with.

6.3. *Actual construction of lower bounds.* Now notice that

(i) in the first problem, the region (5.6.2) includes as well all the F-regions considered under (*i*) of the foregoing subsection (6.2), so that, to the power function P of (6.1.2) we shall have a lower bound given by

$$(6.3.1)\quad P\{\alpha; n_1, n_2, p; \Theta_1, \cdots, \Theta_p\} > 1 - \prod_{i=1}^{p} [1 - P(F \geqq F_0 \text{ or } \leqq F_0' \mid \Theta_i)]$$

(each with $n_1 - 1$ and $n_2 - 1$ degrees of freedom), which is easily calculable.

(ii) in the second problem, the region (5.6.5) includes as well all the F-regions considered under (ii) of the preceding subsection 6.2, so that, to the power function P of (6.1.4) we shall have a lower bound given by

$$(6.3.2)\quad \begin{aligned} &P\{\alpha; n, k, p; \Theta_1, \cdots, \Theta_r\} \\ &\quad > 1 - [1 - P(\text{central } F \geqq F_0)]^{p-r} \prod_{i=1}^{r} [1 - P\,(\text{noncentral } F \geqq F_0 \mid \Theta_i)] \end{aligned}$$

(each with $k - 1$ and $n - k$ degrees of freedom), which is easily calculated; and finally

(iii) in the third problem, the region (5.6.8) includes as well all the | correlation | regions considered under (iii) of the foregoing subsection 6.2, so that, to the power function P of (6.1.6) we shall have a lower bound given by

$$(6.3.3)\quad \begin{aligned} &P\{\alpha; n, p_1, p_2; \Theta_1, \cdots, \Theta_q\} \\ &\qquad > 1 - [1 - P(r^2 \geqq r_0^2 \mid \text{null hypothesis})]^{p_1 - q} \\ &\qquad\qquad \times \prod_{i=1}^{q} [1 - P(r^2 \geqq r_0^2 \mid \rho_i^2 = \Theta_i)], \end{aligned}$$

(each with $n - 2$ degrees of freedom), which is easily calculable, being really the power function of the multiple correlation of the first kind [2], when $p = 2$, for which tables are in part available which could easily be extended with modern computing facilities.

The lower bound (6.3.3) could be easily improved, when $p_2 > p_1$, by the following consideration. Going back to the observations made at the end of subsection 6.2 of this section (on independence between two sets of variates), we notice that since the region (5.6.5) includes $p_1 - 1$ | (*total*) correlation | regions and one (*multiple*) correlation region we shall have a lower bound (easily checked to be larger than (6.3.3)) given by

$$(6.3.4)\quad \begin{aligned} &P\{\alpha; n, p, p_2; \Theta_1, \cdots, \Theta_q\} > 1 - [1 - P(r^2 \geqq r_0^2 \mid \text{null hypothesis})]^{p_1-q} \\ &\times \prod_{i=1}^{q-1} [1 - P(r^2 \geqq r_0^2 \mid \rho_i^2 \equiv \Theta_i)] \times [1 - P(R^2 \geqq r_0^2 \mid \rho_q^2 \equiv \Theta_q)], \end{aligned}$$

where all factors except the last are on | total correlations | distributed with $(n - 2)$ degrees of freedom, while the last factor is on a multiple correlation distributed with $(n - 2)$ degrees of freedom and $(p_2 - p_1)$.

It may be noted that in (6.3.2) *both* sides of the inequality are "known," that is, computationally accessible when $k = 2$, that is, $q = 1$ and $r = 1$, the left-hand side being just the power function of Hotelling's T, while the right hand is also easily available (in this as in all other cases).

7. Concluding remarks. It is of considerable importance at this stage to ask how "good" the lower bounds indicated in (6.3.1), (6.3.2) and (6.3.3) or (6.3.4) are. A lower bound to the power could be said to be "good" if it were (i) close to the actual power, and/or (ii) if it were itself pretty large, being greater than the level of significance α for reasonably large values of the deviation parameters and possibly getting larger as those parameters increase. For all the three tests condition (ii) has been numerically checked to be true over a fairly wide range of test values of the several parameters involved, and part of that material will be offered in a later paper. With regard to condition (i), *in general*, that is, *for small samples*, not only do we not know the actual power (in which case the search for a lower bound would have been redundant) but at the moment we do not even know an upper bound of the expression: (actual power − given lower bound to it) ÷ actual power. In large samples, however, the situation improves and it turns out that the relative error "small," so that the given lower bounds are "good" also in the sense (i).

The next pertinent question now under investigation is whether the proposed test regions (5.6.2), (5.6.5) and (5.6.8) are (a) unbiased and (b) admissible against all relevant alternatives under the respective situations.

Also under investigation is the question as to how these tests compare with the corresponding likelihood ratio or Type II tests. On this it may be observed here, that, except in the degenerate cases where the two methods lead to the identical test, as, for example, the case $k = 2$ under (ii) where both lead to

Hotelling's T, the likelihood ratio tests have a far more difficult small sample (null) distribution problem to contend with than the proposed test. This is with regard to direct usability of the test. The small sample (nonnull) distribution problem (connected with the question of power) would be quite difficult for both types of test, but more so for the likelihood ratio test than for the other. This rules out direct evaluation of power for both types of test, but, while we have fairly good lower bounds to the power of the three different tests proposed, we do not at the moment know of any such lower bounds to the power of the corresponding likelihood ratio tests.

REFERENCES

[The references given below are by no means exhaustive, being merely such as would just suffice for an understanding of this paper.]

[1] R. C. Bose and S. N. Roy, "The distribution of the studentized D^2-statistic," *Sankhyā*, Vol. 4 (1938), pp. 10–38.

[2] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc. London*, Ser. A, Vol. 121 (1928), pp. 654–673.

[3] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[4] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203–224.

[5] H. Hotelling, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[6] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[7] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[8] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 495–516.

[9] D. N. Nanda, "Distribution of a root of a determinantal equation," *Ann. Math Stat.*, Vol. 19 (1948), pp. 47–57.

[10] S. N. Roy, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 4 (1939), pp. 381–396.

[11] S. N. Roy, "The sampling distribution of p-statistics and certain allied statistics on the nonnull hypothesis," *Sankhyā*, Vol. 6 (1942), pp. 15–34.

[12] S. N. Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis," *Sankhyā*, Vol. 7 (1945), pp. 133–158.

[13] S. N. Roy, "Notes on testing of composite hypotheses-II," *Sankhyā*, Vol. 9 (1948), pp. 19–38.

[14] S. N. Roy, "Univariate and multivariate analysis as problems in testing of composite hypotheses-I," *Sankhya*, Vol. 10 (1950), pp. 29–80.

[15] S. N. Roy, "Lecture notes on multivariate analysis," (unpublished).