ON THE MONOTONIC CHARACTER OF THE POWER OF A CERTAIN TEST IN MULTIVARIATE ANALYSIS OF VARIANCE.[1]

By

S. N. Roy

Institute of Statistics

University of North Carolina

Institute of Statistics
Mimeograph Series No. 74

August, 1953

1. This research was done in part under a grant from the Ford Foundation

ON THE MONOTONIC CHARACTER OF THE POWER OF A CERTAIN TEST IN MULTIVARIATE ANALYSIS OF VARIANCE.[1]

By

S. N. Roy

Institute of Statistics

University of North Carolina

Institute of Statistics
Mimeograph Series No.

August, 1953

1. This research was done in part under a grant from the Ford Foundation

1. Summary. In a previous paper [1] three different tests were offered for three types of hypotheses in multivariate analysis, namely, (i) that of equality of the dispersion matrices of two p-variate normal populations, (ii) that of equality of the p-dimensional mean vectors for k p-variate normal populations (which is mathematically identical with the general problem of multivariate analysis of variance of means) and (iii) that of a p-set and a q-set ($p \leq q$) of variates in a (p + q)-variate normal population. Some properties of such tests (and their power functions) were stated and proved, and some further properties of their powers were stated without proof. Among these latter is the property, proved in this paper, that for problem (ii) the power of the test is a monotonically increasing function of the absolute value of each of the corresponding "deviation" parameters, that is, a set of parameters (characterizing the deviation from the null hypothesis) which alone occurs in the power of the test. This shows, incidentally, that the test is an unbiassed one in the sense of Neyman. In proving these, use is made of an unpublished theorem (and its corollaries) due to S. Moriguti, a former student of the author, which, together with Moriguti's own proof, is given in Section 3 of this paper.

2. Introduction. Let us consider the test for (ii), i.e., for analysis of variance, offered in [1]. The notation and terminology to be used here will be the same as in [1] except that X(p x q), for example, will denote that the matrix X consists of p rows and q columns.

2.1 Multivariate analysis of variance test. Let $\underline{\xi}_r$ (p x 1) (r = 1,2,..., k) be the mean vectors of k p-variate normal populations with the same dispersion matrix Σ (assumed to be p.d.), from which random samples of sizes n_r (r = 1, ..., k) are supposed to have been drawn. The null hypothesis H_0 (to be tested) is that $\underline{\xi}_1 = \ldots = \underline{\xi}_k$. Let S^* be the usual sample "between" dispersion matrix of means (weighted by n_r's) and S, the usual sample "within" dispersion matrix (pooled from the dispersion matrices of the k samples). Then it is well known that, almost everywhere, S is p.d. and S^* is at least p.s.d. of rank $q = \min(p, k-1)$, so that, a.e., q of the characteristic roots of S^*S^{-1} are positive and the rest, i.e., p-q are zero. Let θ_q be the largest root. Then the critical region of the test at a level of significance, say α, is:

$$(2.1.1) \qquad \theta_q \geq c,$$

where c is given by

$$P(\theta_q \geq c \mid H_0) = \alpha,$$

so that $c = \theta_\alpha(p, k-1, n-k) = \theta_\alpha$ (say) $\qquad (n = \sum_{r=1}^{k} n_r)$.

Thus c or θ_α is the upper α-point of the distribution of θ_q and depends on α, p, k-1 and n-k.

For a given c and for arbitrary non-null $\underline{\lambda}(p \times 1)$'s, it was shown in [1] that this critical region (2.1.1) was the same as

$$(2.1.2) \qquad \bigcup_{\underline{\lambda}} [\underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda} \geq c] \quad \text{or}$$

$$[\operatorname{Sup}_{\underline{\lambda}} \underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda}] \geq c.$$

The region of acceptance, i.e., $\theta_q < c$, was thus the same as

$$(2.1.3) \qquad \bigcap_{\underline{\lambda}} [\underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda} < c] \quad \text{or}$$

$$[\operatorname{Sup}_{\underline{\lambda}} \underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda}] < c.$$

If we denote by H the usual non-null hypothesis and by Σ^* the usual weighted "between" dispersion matrix of means for the k populations, it is well known and has also been shown in [1] that the power of the critical region, i.e., $P(\theta_q \geq \theta_\alpha \mid H)$ is a function of just the characteristic roots Θ_i's (all non-negative) of the matrix $\Sigma^*\Sigma^{-1}$, aside of course from the other parameters α, p, k-1 and n-k. A lower bound to this power was obtained in [1]. It is shown in the present paper that for a given c, i.e., α, this power is a monotonically increasing function of each of the population characteristic roots, which incidentally proves that the proposed test is unbiassed.

§ 2.2 <u>Canonical form for § 2.1</u>. As is well known and also shown in [1] and [2] we can, for the purpose of discussing the problem of the power function under § 2.1, start, without any loss of generality, from the canonical probability law:

$$\text{(2.2.1)}\quad \text{Constant} \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{k-1} x_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{n-k} y_{ij}^2\right]$$

$$\times \prod_{i=1}^{p}\prod_{j=1}^{k-1} dx_{ij} \prod_{i=1}^{p}\prod_{j=1}^{n-k} dy_{ij},$$

all the variates going from $-\infty$ to ∞. Also the sample "between" dispersion matrix (of means) S^* of § 2.1 is now expressed in the canonical form:

$$\text{(2.2.2)}\quad (k-1)(S^*)_{ii'} \ (i \leq i' = 1, 2, \ldots, p;\ s_{ii'} = s_{i'i})$$

$$= x_{i1}x_{i'1} + \ldots + (x_{ii} + \sqrt{\Theta_i})x_{i'i} + \ldots + x_{ii'}(x_{i'i'} + \sqrt{\Theta_{i'}})$$

$$+ \ldots + x_{i,k-1}x_{i',k-1},$$

where Θ_i's $(i = 1, 2, \ldots, p)$ stand for the p characteristic roots of $\Sigma^*\Sigma^{-1}$ and where Σ is supposed to be p.d. and Σ^* is at least p.s.d. of rank r (say) $\leq \min(p, k-1)$. It is well-known that in this situation, out of the p Θ_i's, p-r are zero and r are positive. Without any loss of generality these positive roots might be taken to be $\Theta_1, \Theta_2, \ldots, \Theta_r$, no ordering in magnitude being necessarily implied. This means that by varying Σ^*, i.e., by considering suitable non-null hypotheses, we could have at the most $q(= \min(p, k-1))$ positive roots. The sample "within" dispersion matrix S of 2.1 will of course be given by

$$(2.2.3) \qquad (n-k)(S)_{ii'} \; (i, i' = 1, 2, \ldots, p) = \sum_{i,i'=1}^{p} \sum_{j=1}^{n-k} y_{ij} \; y_{i'j}$$

Thus, from (2.11) - (2.1.3) we note that the power of the critical region (2.1.1) is given by

$$(2.2.4) \quad P(\theta_q \geq c \mid H) = \text{Integral of (2.2.1) over the domain}$$

$$U_{\underline{\lambda}} \left[\underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda} \geq \theta_\alpha \right] \qquad \text{or}$$

$$\left[\operatorname{Sup}_{\underline{\lambda}} \underline{\lambda}'S^*\underline{\lambda}/\underline{\lambda}'S\underline{\lambda} \right] \geq \theta_\alpha$$

The second kind of error is of course given by the opposite inequality throughout (2.3.4), the U being replaced by $\bigcap$. It should be observed that the constant factor in (2.3.1) does not involve the Θ_i's.

3. Moriguti's results.

(3.1) Main theorem. Let t, u, v and w be independent random variables. Assume that t has a symmetrical distribution with the p.d.f. $\phi(t)$ which is a continuous, decreasing function of $|t|$. Assume further that $P(u > 0) = 1$. Let a be a parameter which does not enter into the distribution of t, u, v and w. Then for the composite random variable $T = [(t + au)^2 + v^2] / w^2$, $P(T \leq c)$ is a decreasing function of $|a|$ for any fixed positive value of c.

Proof. Let the c.d.f. of u, v^2 and w^2 be denoted by $F(u)$, $G(v^2)$, $H(w^2)$, respectively. Then

$$(3.2) \quad P\left\{(t + au)^2 \leq x\right\} = \int_0^\infty dF(u) \int_{-\sqrt{x} - au}^{\sqrt{x} - au} \phi(t)\, dt, \quad (x > 0).$$

This is an even function of a for fixed x, because of the assumed symmetry of $\phi(t)$. We differentiate (3.1) with regard to a and obtain

$$(3.3) \quad \frac{\partial}{\partial a} P\left\{(t + au)^2 \leq x\right\} = \int_0^\infty dF(u) \left[-u\phi(\sqrt{x} - au) + u\phi(-\sqrt{x} - au)\right]$$

$$(x > 0).$$

This is negative for every $a > 0$ and $u > 0$, on account of the assumptions on $\phi(t)$. Hence (3.2) is a decreasing function of $|a|$.

Next, we have

$$(3.4) \quad P(T \leq c) = \int_0^\infty dH(w^2) \int_0^{cw^2} P\left\{(t + au)^2 \leq cw^2 - v^2\right\} dG/v^2),$$

$$(c > 0).$$

The monotonic character with respect to $|a|$ is preserved under the operation. This completes the proof.

<u>Remark</u>. The assumptions can be weakened. For instance, the continuity of $\phi(t)$ is not necessary. Also, v^2 and w^2 can be dependent. The theorem as stated will suffice for the following simple applications.

<u>Some simple applications</u>. Under the assumption of normality of the population, the sample variance ratio with non-central mean-square in the numerator can be written in the form $F = (n_2/n_1)\left[(t + a)^2 + \chi^2_{n_1-1}\right] / \chi^2_{n_2}$, when t is a $N(0, 1)$, χ^2's are random variables each distributed as chi-square with the number of degrees of freedom indicated by the subscript, and (These) three are independent. For a p-variate normal population, the sample multiple correlation coefficient R and the sample partial correlation coefficient r can be brought into the forms:

$$R^2/(1 - R^2) = \left[(t + a\chi_{n-1})^2 + \chi^2_{p-2}\right] / \chi^2_{n-p} \quad \text{and}$$

$$r^2/(1 - r^2) = (t + a\chi_{n-p-1})^2 / \chi^2_{n-p}, \quad \text{respectively,}$$

where the symbols have similar meanings as in F and where a is equal to $\rho/(1 - \rho^2)^{1/2}$, ρ being the corresponding population correlation coefficient. The ordinary, i.e., total correlation coefficient is a special case where $p = 2$. This implies that the powers of the respective critical regions: $F \geq F_0$, $R^2 \geq R_0^2$ and $r^2 \geq r_0^2$, are each an increasing function of the corresponding "non-centrality parameter" $|a|$.

4. Some further observations on Moriguti's theorem and its uses. It will be seen on a close examination of the structure of (3.3) and its tie-up with (3.2) and (3.4) that the secret of the proof lies in our being able to get in (3.3), after the t-integration, a negative factor in the integrand which leaves the whole integrand negative for all $a > 0$ such that the negativeness is not upset in the subsequent stages of integration. Under the assumptions on the p.d.f. of t, this negative factor would arise out of the t-integration whenever the modulus of the upper limit of the t-integration is less than that of the lower limit of the integration. A formal generalization of the theorem (3.1) with weaker (sufficient) conditions could be given but that would not be necessary for the limited purposes of this paper.

5. Monotonic character of the power of the multivariate analysis of variance test. Going back to § 2.3 consider the complement of the power, i.e., the second kind of error, of the test, which will be the integral of (2.2.1) over the domain

$$(5.1) \qquad \bigcap_{\underline{\lambda}} \left[\underline{\lambda}'S^*\underline{\lambda} \,/\, \underline{\lambda}'S\underline{\lambda} < \theta_\alpha \right] \quad \text{or}$$

$$\left[\operatorname{Sup}_{\underline{\lambda}} \underline{\lambda}'S^*\underline{\lambda} \,/\, \underline{\lambda}'S\underline{\lambda} \right] < \theta_\alpha .$$

It will suffice to show the monotonic character of this integral with respect to variation of any one parameter, say $\sqrt{\Theta_1}$. To this end, remembering that $\underline{\lambda}'$ is the row vector $(\lambda_1, \lambda_2, \ldots, \lambda_p)$, we might, without any loss of generality, put $\lambda_1 = 1$ and rewrite (5.1) as

$$(5.2) \qquad \bigcap_{\underline{\lambda}} \left[\left\{ (x_{11} + \sqrt{\Theta_1}) + \lambda_2 x_{21} + \ldots + \lambda_p x_{p1} \right\}^2 \right.$$

$$\left. + \sum_{j=2}^{k-1} \left\{ \sum_{i=1}^{p} \lambda_i (x_{ij} + \delta_{ij}) \right\}^2 < \theta_\alpha \sum_{j=1}^{n-k} \left(\sum_{i=1}^{p} \lambda_i y_{ij} \right)^2 \right] ,$$

where $\delta_{ij} = 0$ if $i \neq j$ and $= \sqrt{\Theta_i}$ if $i = j$, and where $\lambda_i = 1$. Consider now the integral of (2.2.1) over the domain (5.2) and, for given values of the other v riates, perform first the integration over x_{11}. The contribution to the total p.d.f. (2.2.1) made by x_{11} is just the factor const. $\exp[-\frac{1}{2}x_{11}^2]$ which satisfies the conditions of theorem (3.1). The upper and lower limits of this x_{11}-integration are ℓ_1 and ℓ_2 given by

$$(5.3) \qquad \ell_1 = \operatorname{Inf}_{\underline{\lambda}}\Big[-\sqrt{\Theta_1} - \Big(\sum_{i=2}^{p}\lambda_i x_{i1}\Big) + \Big\{\Theta_\alpha \sum_{j=1}^{n-k}\Big(\sum_{i=1}^{p}\lambda_i y_{ij}\Big)^2 - \sum_{j=2}^{k-1}\Big(\sum_{i=1}^{p}\lambda_i(x_{ij} + \delta_{ij})\Big)^2\Big\}^{1/2}\Big] \quad \text{and}$$

$$\ell_2 = \operatorname{Sup}_{\underline{\lambda}}\Big[-\sqrt{\Theta_1} - \Big(\sum_{i=2}^{p}\lambda_i x_{i1}\Big) - \Big\{\Theta_\alpha \sum_{j=1}^{n-k}\Big(\sum_{i=1}^{p}\lambda_i y_{ij}\Big)^2 - \sum_{j=2}^{k-1}\Big(\sum_{i=1}^{p}\lambda_i(x_{ij} + \delta_{ij})\Big)^2\Big\}^{1/2}\Big]$$

It is easy to check that the x_{11}-integral is an even function of $\sqrt{\Theta_1}$ and also that, for all positive $\sqrt{\Theta_1}$, $|\ell_1| \leq |\ell_2|$ and, a.e., $|\ell_1| < |\ell_2|$. If we now differentiate, with regard to $\sqrt{\Theta_1}$ the integral of (2.3.1) over the domain (5.2) we immediately obtain, through the x_{11}-integral, as in (3.3) an integrand which is

$$(5.4) \qquad -\exp\Big[-\tfrac{1}{2}\ell_1^2\Big] + \exp\Big[-\tfrac{1}{2}\ell_2^2\Big],$$

and for all positive $\sqrt{\Theta_1}$ this is, a.e., negative, since, a.e., $|\ell_1| < |\ell_2|$.

Notice that (5.4) is free from all $\underline{\lambda}$'s and is a pure function of Θ_i's ($i = 1, 2, \ldots, p$) and all stochastic variates other than x_{11}. If now we integrate out over the other stochastic variates it is easy to see that, since for all positive $\sqrt{\Theta_1}$ the integrand is, a.e., negative, therefore the integral will also be negative for all positive $\sqrt{\Theta_1}$. This integral being just the differential coefficient with regard to $\sqrt{\Theta_1}$ of the integral of (2.2.1) over (5.2), it follows that, for all positive $\sqrt{\Theta_1}$, the latter integral (which is also an even function of $\sqrt{\Theta_1}$) is a decreasing function of $\left|\sqrt{\Theta_1}\right|$. Thus the second kind of error of the test (2.1.1) is a decreasing and, therefore, the power of the test is an increasing function of each $\left|\sqrt{\Theta_i}\right|$ separately.

We have shown that the power of the test is an increasing function of each of the parameters $\Theta_1, \Theta_2, \ldots, \Theta_q$ separately, i.e., under variation of any one of them, the others being held fixed. Remembering that the Θ_i's are the characteristic roots of the matrix $\Sigma^* \Sigma^{-1}$ one can easily show further that corresponding to two alternatives H_1 and H_2 we should have

(5.5) $$P(\theta_q \geq c \mid H_1) > P(\theta_q \geq c \mid H_2)$$ if ~~and only if~~

$$\sum_{i=1}^{q} \Theta_i^{(1)} > \sum_{i=1}^{q} \Theta_i^{(2)}, \quad \sum_{i \neq j=1}^{q} \Theta_i^{(1)} \Theta_j^{(1)} > \sum_{i \neq j=1}^{q} \Theta_i^{(2)} \Theta_j^{(2)},$$

$$\ldots, \prod_{i=1}^{q} \Theta_i^{(1)} > \prod_{i=1}^{q} \Theta_i^{(2)}$$, or in other words, if ~~and only if~~

$$\mathrm{tr}_1\, \Sigma_1^* \Sigma_1^{-1} > \mathrm{tr}_1 \Sigma_2^* \Sigma_2^{-1}\ , \quad \mathrm{tr}_2 \Sigma_1^* \Sigma_1^{-1} > \mathrm{tr}_2 \Sigma_2^* \Sigma_2^{-1}\ , \ \ldots, \ \mathrm{tr}_q \Sigma_1^* \Sigma_1^{-1} > \mathrm{tr}_q\, \Sigma_2^* \Sigma_2^{-1}\ .$$

Here Σ_1^*, Σ_1 and $\Theta_i^{(1)}$'s correspond to H_1 and Σ_2^*, Σ_2 and $\Theta_i^{(2)}$'s to H_2.

6. Concluding remarks. By using the same method (discussed in sections 3 and 5) it is possible to show that the power of the test of independence between two sets of variates (or, in other words, the test for problem (iii)) given in [1] is also a monotonically increasing function of each of the 'deviation parameters', i.e., of the population canonical correlation coefficients. This will be given in a later paper.

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. 220-238.

2. S. N. Roy, "Lecture notes on multivariate analysis", (Unpublished).