UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

Mathematical Sciences Directorate
Air Force Office of Scientific Research

SOME REMARKS ON NORMAL MULTIVARIATE ANALYSIS OF VARIANCE

by

S. N. Roy

August, 1961

Contract No. AF 49(638)-213

This paper brings together certain developments in normal multivariate analysis of variance in recent years that are based on a particular approach in which special stress is laid on decomposition of a complex question into more elementary components, and both the inference procedure and the design pattern are aligned to the particular decomposition used. While some of the developments are available in print and the rest still unpublished, this paper discusses in greater detail than heretofore the general philosophy and its role in the actual developments. The problems are also viewed against the wider background of multifactor multiresponse experiments (where some of the factors may be structured and some unstructured and similarly for the responses). Among the facets considered is one which shows that MANOVA (multivariate analysis of variance) is, in a sense, more general, and, in another sense, more restricted than ANOVA (univariate analysis of variance). The paper being primarily concerned with general considerations, the bulk of the details of various derivations is left out.

Institute of Statistics
Mimeograph Series No. 304

SOME REMARKS ON
NORMAL MULTIVARIATE ANALYSIS OF VARIANCE *

by

S. N. Roy
Department of Statistics
University of North Carolina

1. Introduction. It is profitable to look at normal multivariate analysis of variance from the wider perspective of multifactor multiresponse experiments discussed in [1, 11]. Factors are what we can either control or watch (and, with some justification, assume to be nonstochastic, in the sense of concomitant variables), and a response type is what we are looking for. Under any design and sampling scheme what we get in the first instance is a set of experimental units, and, according as we study each experimental unit (subjected to a suitable factor-level combination) with respect to one or more response types or characteristics, we have a univariate or a multivariate problem. The relationship between the factors and the responses (or rather the response types) can be described as, in some sense, a generalization of the classical cause and effect relationship. In these situations the response types or responses are supposed to have a multivariate probability distribution whose nature or characterization is supposed to depend on the factors (or factor-level combination), and one of the main purposes of experimental designs and analysis of variance is to study, as effectively as our resources would permit, the nature of this dependence. In the classical analysis of variance, the mean, bias, variance, least squares, etc. are merely tools for the study of this dependence under a somewhat restricted model, and it may be, and, in fact, has been found necessary under other

*This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

models to consider a different set of tools which would play under these models the same role that the classical tools play under the classical model. The different factors (each with its own set of levels) are what are usually called assignable sources of variation, and a conscious attempt is made, as part of the planning, to pinpoint as many of these as possible, and then to regard the other possible sources of variation (unknown and inaccessible to us and possibly less important in their contribution to the variation) as being lumped together into something that produces the error, or, in other words, something that transforms into a probability picture what would otherwise have been a functional relationship. Taking any response (or response type) we observe that it can be (i) purely categorical, or (ii) categorical with an implied ranking or (iii) discrete or, finally, (iv) continuous. In the first case we shall say that the response is unstructured and in the last three cases that the response is structured. For any factor (with its set of levels) there is a similar division into four classes with a similar characterization as an unstructured or a structured factor. With regard to the factor-response relationship or dependence, consider, for simplicity of discussion, one factor and one response in which case we have the four possibilities: (a) both unstructured, (b) the factor unstructured but the response structured, (c) the factor structured but the response unstructured, and, finally, (d) both structured. Toward the study of the dependence or relationship just mentioned, there are certain questions that are relevant to the case (a). We can ask the same questions in the case of (b) or(c), but they would no longer be as useful or as relevant. Some other questions,which it is not possible to pose in the case of (a), will be now far more relevant and important. We have an exactly similar situation in the passage from (b) or (c) to the case (d). In many complex experiments (especially physico-chemical or industrial or

biological at a certain level) we have a multifactor, multiresponse situation in which some of the factors are of type (i), some of type (ii), some of (iii) and some of (iv), with a similar division into four classes of the responses. The questions that are interesting for and relevant to the total setup are necessarily somewhat complex, each component of the total relationship or dependence, based on one response and one factor (after a certain transformation), having its own characteristic kind of questions according as the pair is of the type (a), (b), (c) or (d). In most experimental situations there is at least one factor (generally unstructured and with levels 1, 2, 3, etc., the numbers, for the unstructured case, being necessarily merely nominal) in which we are not ordinarily interested but which, nevertheless, enters into the experimental scheme as an unavoidable nuisance. This factor is usually called a set of blocks, and while we are ordinarily interested in the relationship between the responses and the other factors it is found that in the study of this relationship we are much worse off if we ignore this factor than if we properly plan the experiment with due regard to this factor and then study the dependence in which we are interested after eliminating the effect of this factor. This is a remarkable achievement and this and other important aspects of the general philosophy (that consists in asking nature a set of simultaneous questions rather than one question at a time) and of the general planning (characterized by replication, randomization, local control, etc.) are set forth in the pioneering book by Fisher [3], now a classic, followed by a whole host of books in recent years, some of them quite good, among which, for illustration, mention may be made of [2, 5]. It is, therefore, unnecessary to discuss these aspects here.

By and large, the classical analysis of variance deals with one response (even when several responses are considered they are, most of the time, treated separately) which is continuous (either from the start or after a

transformation that is sometimes questionable), with a mean and a variance, and for which a normal distribution is assumed for some, though by no means for all purposes. So far as the factors are concerned it is a multifactor setup consisting of (1) two unstructured factors, e.g., a set of blocks and a set of treatments like several varieties of corn or (2) two or more structured factors like temperature, pressure, etc. with or without an unstructured factor like a set of blocks or (3) a mixture of (1) and (2), consisting of some unstructured and some structured factors. The present discussion is concerned with the case where the factor setup is the same as in the classical treatment but where we have several continuous responses (to be treated jointly, not separately) which follow a multivariate distribution whose means and dispersion matrix exist and for which normality is assumed for certain, but by no means for all purposes. The more general situation involving other response types has been partially discussed in other papers and will not be considered. Even under the narrower assumptions of this paper the discussion will be mostly restricted to the still narrower model known as model I, and attention will be drawn, in particular, to certain aspects of the development implicit but not brought out very prominently in the previous publications by the author and his group. Model II and mixed models of normal multivariate variate analysis or MANOVA, considered in [10, 12], will not be discussed here. Suffice it to say, however, that those aspects of Model I, not discussed in any detail in previous publications, but particularly stressed here, will carry over to Model II and mixed models of MANOVA as well, and will not, therefore, be separately considered here.

2. Model I and the associated analysis.

2.1. The linear model I. $\underset{pxl}{X_i}$ $(i = 1,2,\ldots,n)$ is a set of n p-dimensional vectors, each having the same dispersion matrix $\underset{pxp}{\Sigma}$ and the correlation between any component of X_i and any of $X_{i'}$ $(i \neq i' = 1,2,\ldots,n)$ being zero. In symbols, if we roll out the set of vectors as a pn dimensional vector $\underset{pnxl}{X^*}$, then this is supposed to have the dispersion matrix $\underset{pxp}{\Sigma} \times \underset{nxn}{I}$, where the middle symbol stands for a Kronecker product. Furthermore, $\underset{pxn}{X} \equiv [X_1, X_2, \ldots, X_n]$ is assumed to have an expectation given by

$$(1) \quad \underset{nxp}{\mathcal{E}(X')} = \underset{nxm}{A}\ \underset{mxp}{\xi},$$

where A, to be called the model matrix, is supposed to be a known matrix of constants that might also include the observed concomitant variables given partly by the design and partly by what the experimental statisticians call the "model", and ξ is a matrix of unknown parameters (necessarily nonstochastic). A little reflexion will show that the design and what the experimental statisticians call the "model" are both directly associated with the experimental units, and hence the matrix A is the same no matter whether the problem is a univariate or a multivariate one, i.e., whether p = 1 or > 1, i.e., whether X' and ξ are n and m dimensional vectors or nxp and mxp matrices. We assume, in addition that rank (A) = $r \leq m \leq n-p$. The inequality connecting m and n-p might be easily relaxed or done away with, but, without any essential loss of generality, it can also be retained as such. The inequality $r \leq n-p$ is, however, essential, at least for the kind of treatment we are going to consider here. Thus, in general, what we need is that $r \leq \min(m, n-p)$. The problem of point estimation of linear functions of $\underset{n}{\xi}$ under this model has been discussed in [9], and may be skipped here. We pass on directly to the problem of testing and confidence

interval estimation. For such problems we assume, furthermore, that $\underset{pxn}{X} \equiv [\chi_{1_1}, \ldots, \chi_{1_n}]_p$ are independently distributed p vectors such that $\chi_i (i=1,2,\ldots,n)$ is $N [\underset{pxl}{\xi}(\chi_i), \underset{pxp}{\Sigma}]$.

2.2. Linear hypothesis under the linear model I. Rolling $\underset{mxp}{\zeta}$ out into a vector $\underset{mpxl}{\zeta^*}$, and in keeping with the practice in univariate analysis of variance (ANOVA) we might be, quite legitimately, interested in the linear hypothesis expressed in the form.

$$(2) \qquad \underset{s^*xmp}{C^*} \; \underset{mpxl}{\zeta^*} = \underset{s^*xl}{0} \quad ,$$

where C^* is a given matrix of rank $s^* \leq mp$. But, unfortunately, unlike the ANOVA, it is not possible to carry out (free from the nuisance parameters Σ) a test of this hypothesis in the MANOVA, the obstacle probably lying in the fact that here we have a (pxp) symmetric matrix Σ to contend with unlike the ANOVA where we have only a σ^2 to deal with. Thus we have to be satisfied with something less which, however, is still sufficiently general to serve most of our purposes. The more restricted type of linear hypothesis can be expressed in the form

$$(3) \qquad H_o : \underset{sxm}{C} \; \underset{mxp}{\zeta} \; \underset{pxu}{U} = \underset{pxu}{0} \quad ,$$

where C and U are given matrices of constants (to be called a pair of hypothesis matrices) of ranks $s(\leq r)$ and $u(\leq p)$ respectively. For testability (in the strong sense) of this H_o against the general alternative

$$(4) \qquad H: \; C \; \zeta \; U \neq 0 \quad (= \underset{sxu}{\eta}, \text{ say}),$$

a necessary and sufficient condition [9,13] is

$$(5) \qquad \text{rank} \underset{m}{[\underset{C}{A}]}_s^n = \text{rank} \underset{nxm}{(A)} \quad .$$

It is easy to see that (3) is a special case of (2), and, as pointed out just now, this is what we have to be satisfied with in MANOVA. Specializing C

(subject to (5)) we can handle all types of treatment contrasts (as in ANOVA), both when the set of treatments is unstructured (as in several varieties of wheat, for example) or when the different treatment combinations form a set of structured factors as in classical analysis of factorial experiments (in terms of temperature, pressure combinations, for example). U is many problems will be just the unit matrix I, but by properly specializing U we can handle problems of profile analysis, or of comparison of growth curves or growth surfaces under different treatments (where each treatment might be a complex combination), if necessary even after elimination of "block effects". In short, C relates the different treatments among themselves (mostly in terms of contrasts) while U relates the different variates (or responses) among themselves (again, mostly in terms of contrasts). U, of course, is peculiar to MANOVA and does not have any analogue in ANOVA.

An equivalent form [9,13] of H_o is

$$(6) \qquad H_o: \underset{mxp}{\zeta}\ \underset{pxu}{U} = \underset{mxk}{B}\ \underset{kxu}{\Theta},$$

where B is a matrix of known constants, Θ is a set of unknown parameters; the structure of B (in terms of dimensionality and rank) has a certain relation to that of C and A, when we pass from (3) to (6), that is explained in [9,13]. The form of the alternative that parallels (4) is obvious and need not be separately indicated. As in case of form (3), in the case of form (6), B and U are called a pair of hypothesis matrices. For testability (in the strong sense) the condition in the case of forms (6) is [9,13].

$$(7) \qquad 0 < \text{rank } A - \text{rank } AB = m - \text{rank } B .$$

Although, as mentioned above, (3) and (6) are equivalent in that one can pass from one to the other, yet there is a wide range of problems where (3) is a natural and convenient form as in most problems of ANOVA and MANOVA (involving

unstructured or structured factors and including univariate or multivariate analysis of covariance and regression). However, there are other problems, as, for example, testing linearity of regression (against a general alternative) where (6) is a natural convenient form. Thus it is convenient to have an explicit formulation in terms, of (6) (subject to (7)).
We shall assume for purposes of the present discussion that (5) and (7) hold for (3) and (6) respectively. We note that for either form (3) or (6), putting u = 1 (or, if U is the identity matrix as in a wide class of problems, then putting p=1) we are back in the univariate case for which we have the customary F-test.

No matter whether H_o is expressed in the form (3) or (6) in the build-up of any test procedure of a certain class that parallels the F-test of ANOVA, two nxn symmetric positive semi-definite matrices play a key role. They are called respectively the "matrix due to the error" and the "matrix due to the error" and will be denoted here by Q_{error} and Q_{H_o} . Explicit expressions are given in [8,9,12] for Q_{error} in terms of A and for Q_{H_o} in terms of A , and C or B according as H_o is expressed in the form (3) or (6). These matrices of course are the same for MANOVA as for ANOVA, and, in the extensive literature of ANOVA, issued over the last thirty years, these matrices, for most situations considered in ANOVA so far, are implicitly available in forms outwardly somewhat different from those given in [8,9,12]. When MANOVA becomes a little more familiar to the general statistical public the practice will gradually develop of giving, for any type of design and problem of interest, these two matrices, rather than the "s.s. due to the error", "s.s. due to the hypothesis", etc., so that the experimenter can use these matrices to handle the ANOVA or the MANOVA, as the case might be. Based upon these two matrices there are two other matrices (involving also the observed variates or responses) that play a key role,

namely,

$$(8)\qquad (n-r)\underset{uxu}{S_{error}} = \underset{uxp}{[U'} \quad \underset{pxn}{X} \quad \underset{nxn}{Q_{error}} \quad \underset{nxp}{X'} \quad \underset{pxu}{U]} ,$$

$$s\underset{uxu}{S_{H_o}} = \underset{uxp}{[U'} \quad \underset{pxn}{X} \quad \underset{nxn}{Q_{H_o}} \quad \underset{nxp}{X'} \quad \underset{pxu}{U]} .$$

For u = 1, we are back in the univariate case and these will be just the "variance due to the error" and the "variance due to the hypothesis." No matter whether H_o is expressed as (3) or (6), three tests (of a certain class) have been proposed in recent years with critical regions

$$(9)\qquad W_1: \ |S_{error}| \ / \ \left|\frac{1}{n-r+s}[s S_{H_o} + (n-r)S_{error}]\right| \leq \lambda_1 ,$$

$$(10)\qquad W_2: \ tr[S_{H_o} S_{error}^{-1}] \geq \lambda_2$$

and

$$(11)\qquad W_3: \ ch_{max}[S_{H_o} S_{error}^{-1}] \geq \lambda_3 ,$$

where "tr" stands for the trace of the matrix and "ch_{max}" for the largest characteristic root and where λ_1, λ_2 and λ_3 are given by

$$(12)\qquad P[W_1|H_o] = P[W_2|H_o] = P[W_3|H_o] = \alpha ,$$

α being any preassigned level of significance (or size). The associated tests are called respectively the λ-criterion, the sum of the roots test and the largest root test. For u=1 (i.e., the univariate case) each of these three tests (based on W_1, W_2 and W_3 respectively) reduces to the customary F-test. The power functions of these test procedures are extremely hard to obtain in an explicit form, and even if they were so obtained, it still would not serve any useful purpose. However, even without explicit expressions there are other means of showing that in each case the power function depends, aside from the d.f. u, s, n-r (and of course the level of significance α) on the u roots $o \leq \zeta_1 \leq \zeta_2 \leq \ldots \leq \zeta_u < \alpha$ (some of which might be

zero) of the symmetric positive semi-definite matrix

$$(13)\qquad \underset{uxs}{[\eta'} \quad \underset{sxs}{Q^*_{H_o}} \quad \underset{sxu}{\eta} \quad (\underset{uxp}{U'} \quad \underset{pxp}{\Sigma} \quad \underset{pxu}{U})^{-1} \underset{}{]} ,$$

where $Q^*_{H_o}$ is a sxs symmetric matrix given in [7,8,12] in terms of A, and C or B according as H_o is expressed as (3) or (6). Let us write the three power functions symbolically as

$$(14)\qquad \text{(i)}\ \psi_1(\alpha, u, s, n\text{-}r, \zeta_1, \ldots, \zeta_u) ,$$

$$\text{(ii)}\ \psi_2(\alpha, u, s, n\text{-}r, \zeta_1, \ldots, \zeta_u) ,$$

$$\text{(iii)}\ \psi_3(\alpha, u, s, n\text{-}r, \zeta_1, \ldots, \zeta_u).$$

It has been proven that (a) ψ_3 has a very simple lower bound which is a monotonically increasing function of each $\zeta_1, \zeta_2, \ldots, \zeta_u$, separately [7,8], and (b) ψ_3 itself monotomically increases with each $\zeta_1, \ldots, \zeta_u$, separately [14]. The result (b) has also been proven for ψ_2 and ψ_1, though not yet published. Also, "inverting" a probability statement based on the distribution associated with (11) it has been possible in recent years to put, with a conservative confidence coefficient, simultaneous confidence bounds on the largest root of the "total" $[\eta' \quad Q^*_{H_o} \quad \eta]$ and of the "partials", where largest root of the "total" and of the "partials" can be interpreted as measures of deviation from H_o (and of its components) in the directions of the alternative to H (and of the corresponding alternatives to its components). This has been described in detail in another paper submitted to the current session of the ISI, where a bibliography is also supplied. It will not, therefore, be discussed in this paper. As far as I am aware it has not yet been possible to "invert" (9) or (10), in the same sense, and obtain confidence bounds on similar parametric functions, although I believe that it may not be impossible to eventually accomplish this with (9). For a significantly

different approach to MANOVA based on a slip-down procedure and the union-intersection of a finite number of F-regions see [6] .

3. The union-intersection principle and its impact on MANOVA and design of experiments [7,8,15]. The hypothesis H_o and the alternative H (taken conveniently in the form (3)) can be expressed (note that this expression is by no means unique) as

$$(15) \qquad H_o = \bigcap_a [C \underset{uxl}{\zeta} U \underset{sxm}{a} \underset{mxp}{=} 0 \underset{pxu}{]}, \quad H = \bigcup_a [C \underset{uxl}{\zeta} U a \neq 0 \underset{uxl}{]},$$

or

$$H_o = \bigcap_a H_{oa} \quad \text{and} \quad H = \bigcup_a H_a \quad \text{(say)},$$

or, in other words, as the intersection of a set of components H_{oa} and the union of the complementary sets. With H_{oa} and H_a (for any a) we are back in the univariate problem. In this context the a can be interpreted as a system of weights, and, in particular, if U = I, then U in a can be interpreted as a set of weights to the different variates (or responses), on economic or other considerations, so that one has a "total" outburn in terms of that set of weights. Now remembering that if $\underset{uxl}{y}$ is an $N[\underset{uxl}{\mathcal{E}}(y), \underset{uxl}{\Sigma^*}]$, then $\underset{lxu}{y'}\ \underset{uxl}{a}$ is $N[\underset{lxu}{\mathcal{E}}(y)'\ \underset{uxl}{a}, \underset{lxu}{a'}\ \underset{uxu}{\Sigma^*}\ \underset{uxl}{a}]$, and making use of some other simple results [7,8] we have for H_{oa} against H_a, at a level of significance α^*, the customary F-test (with a number of well known strong properties) whose critical region can be expressed as

$$(16) \qquad W_{3a}: \ F_{a(s,n-r)} \equiv \frac{a'S_{H_o}a}{a'S_{error}a} \geq \lambda_3 ,$$

where $P[W_{3a} | H_{oa}] = \alpha^*$, and $F_{a(s,n-r)}$ (depending, among other things, on the particular a) is an F with d.f. s and n-r. The critical region W_3 of (11) is expressible as

$$(17) \qquad W_3 = \bigcup_a W_{3a} = \bigcup_a [F_{a(s,n-r)} \geq \lambda_3),$$

and the acceptance region, which is its complement, as

$$(18) \qquad \bar{W}_3 = \bigcap_a \left[\frac{a' S_{H_o} a}{a' S_{error} a} < \lambda_3 \right] .$$

Thus, if we used the procedure based on (11) we have a test for H_o against H with an acceptance region which is the intersection of the customary acceptance regions for the components H_{oa} against the corresponding alternatives H_a and a rejection (or critical region) which is the union of the customary critical regions for the components. It is to be particularly stressed that having first interpreted a as a set of weights we obtain finally $\bar{W}_3$ and W_3 as weight invariant regions, and the test procedures based thereon as weight-invariant procedures. Notice that $\alpha^* < \alpha$. This means that if we are accepting H_o with a probability $1-\alpha$ (under H_o) we are accepting each component H_{oa} with a (much larger) probability $1-\alpha^*$ (under H_{oa}), while, if we are rejecting H_o at a level α, we are rejecting H_{oa} (for at least one a) at a more stringent level α^* ($< \alpha$), i.e., at least, at the level α. It should be observed that even if we are accepting H_o with a probability $1-\alpha$ (under H_o) we might still be rejecting H_{oa} (for many a's) <u>at the level α</u>. The λ-criterion and the "sum of the roots" test could also be interpreted as examples of the general union-intersection principle but based on other decompositions (different from the one considered here) of H_o (and H); and for this problem those decompositions (and their consequences) will be a lot more difficult to study than for the largest root test, although there could be other situations, where the λ criterion or the "sum of the roots" might turn out to be easier to handle in terms of decomposition and its interpretation. The above is an example of the union-intersection principle in which the total hypothesis H_o is interpreted as a decomposition into components H_{oa} such that each component is, as it were, a pseudo-univariate hypothesis. In a sense, one might call it a variate wise decomposition (a term that would be

even more apt if U were a unit-matrix). Also each component carries with it an F-test. However, each H_{oa} (with its own F_a) can be further interpreted as an intersection. With regard to this part of the discussion, for simplicity, we shall assume a setup with two unstructured factors like blocks and treatments, in any general connected design with only one replication in each cell, the corresponding contribution vectors (in an additive model) being denoted by β_j and τ_j $[j = 1,2,\ldots,b$ (no. of blocks) and $i = 1,2,\ldots,v$ (no. of treatments)$]$. Then the customary hypothesis of "no treatment effect" can be written as

$$(19) \qquad H_o: \underset{pxl}{\tau_1} = \underset{pxl}{\tau_2} = \ldots = \underset{pxl}{\tau_v} \text{ against } H: \text{ not } H_o ,$$

and, given the design, it is easy to write the A and C matrices and check that U = I. In terms of the decomposition, indicated above, into pseudo-univariate components, H_o and H can be expressed as

$$(20) \qquad H_o = \bigcap_a H_{oa} = \bigcap_a [a'\tau_1 = a'\tau_2 = \ldots = a'\tau_v] ,$$

where the intersection is overall nonnull $\underset{pxl}{a}$, and a similar decomposition for H.

However, we can go one step further and regard the H_{oa} (for each a) as itself the intersection of treatment contrasts. In other words denoting by $\underset{pxv}{\tau} = [\underset{1}{\tau_1}, \underset{1}{\tau_2}, \ldots, \underset{1}{\tau_v}]p$, we can express H_o as

$$(21) \qquad H_o \equiv \bigcap_{a,b} H_{o\,a\,b} = \bigcap_{a,b} [\underset{1xp}{a'} \underset{pxv}{\tau} \underset{vxl}{b} = 0] ,$$

over all nonull a and all b's such that

$$(22) \qquad \text{(i) } \underset{1xv}{b'} \underset{vxl}{1} = 0 \text{ and (ii) } b'b = 1,$$

where $1' = \underset{1xp}{[1,1,\ldots,1]}$. The condition (i) indicates a linear contrast. (ii) is merely a normalization condition that is convenient but not essential. Notice that a is related to linear compounds of the variates, i.e., to a system of weights to the different variates, while b is related to treatment contrasts. For any H_{oa} there are, however, (infinitely many) other types of decomposition into H_{oab}, of which let us consider two interesting ones, by way of illustration

$$(23) \qquad H_o = \bigcap_a \bigcap_{i\neq i'} H_{oaii'} = \bigcap_a \bigcap_{i\neq i'=1}^{v} \underset{lxp}{[a'} \underset{pxl}{(\tau_i} - \underset{pxl}{\tau_{i'})} = 0]$$

and

$$(24) \qquad H_o = \bigcap_a \bigcap_i H_{oai} = \bigcap_a \bigcap_{i=1}^{v-1} \underset{lxp}{[a'} \underset{pxl}{(\tau_i} - \tau_v) = 0] \ .$$

So far as the variates or responses are concerned, we have the same pattern as in (21), but so far as the treatments are concerned, (23) is based on pairwise contrasts and (24) is based on contrast of each treatment (from 1 to v-1) with a standard one (say the v^{th} one). For (23) and (24) let us set up the acceptance regions

$$(25) \qquad \bar{W}_3^{(1)} : \bigcap_{i\neq i'=1}^{v} \bigcap_a \underset{lxp}{[a'} \underset{pxl}{(\tau_i} - \underset{pxl}{\tau_{i'})}\Big)^2 / (\underset{lxp}{a'} \Sigma_{nu} a) \leq \mu_3^{(1)}]$$

and

$$(26) \qquad \bar{W}_3^{(2)} : \bigcap_{i=1}^{v-1} \bigcap_a [(a'(\tau_i - \tau_v))^2 / (a' \Sigma_{nu} a) \leq \mu_3^{(2)}]$$

where $\tau_i - \tau_{i'}$ in (25) and $\tau_i - \tau_v$ in (26) are the maximum likelihood estimates of the contrast vectors $\tau_i - \tau_{i'}$ and $\tau_i - \tau_v$, and Σ_{nu} in both cases refers to the "error dispersion matrix" based on the overall residual. Also the constants $\mu_3^{(1)}$ and $\mu_3^{(2)}$ are given by

$$(27) \qquad P[\bar{W}_3^{(1)} | H_o] = P[\bar{W}_3^{(2)} | H_o] = 1 - \alpha$$

It is to be observed now that for a given a, and a given (i,i') for (25), and a given i for (26), the expressions within the square parenthesis in (25) and (26) represent, in each case, a two-sided t-region with appropriate d.f. and with a number of optimal properties. Taking the intersection over a we have in each case a region based on Hotelling's T^2 with appropriate d.f. Let us denote these by $T^2_{ii'}$ and T^2_{iv} for the two cases. Then (25) and (26) can be rewritten as

$$(28) \qquad \tilde{W}_3^{(1)} : \bigcap_{i \neq i'=1}^{v} \left[T^2_{ii'} \leq \mu_3^{(1)} \right]$$

and

$$(29) \qquad \tilde{W}_3^{(2)} : \bigcap_{i=1}^{v} \left[T^2_{iv} \leq \mu_3^{(2)} \right] .$$

Let us denote the powers of the tests based on (21), (28) and (29) by P_3, $P_3^{(1)}$ and $P_3^{(2)}$. Then it can be proven that for a deviation $(\tau_i - \tau_{i'})$ (other deviations being zero and $i \neq i' = 1,2,\ldots,v$), $P_3^{(1)} > P_3$; and for a deviation $(\tau_i - \tau_v)$ (other deviations being zero and $i = 1,2,\ldots,v-1$), $P_3^{(2)} > P_3^{(1)} > P_3$. This means that, for pairwise contrasts, $W_3^{(1)}$ makes a better discrimination than W_3 and also $W_3^{(2)}$ (except for the pair (i,v)), and for contrasts (i, v), against a standard one, $W_3^{(2)}$ is better than $W_3^{(1)}$ and $W_3^{(1)}$ is better than W_3. A total hypothesis, decomposed in any manner, will be called a structured hypothesis. This use of the term "structured" should not be confused with the use involved in a "structured factor or response". Other interesting decompositions are discussed in [15].

The procedure consists essentially in expressing a complex (composite) hypothesis as the intersection of a number of more elementary (composite) components where each such elementary component has a test procedure that has an optimal property in a very natural and meaningful sense. For the complex hypothesis a test procedure is suggested that has an acceptance

region which is the intersection of the acceptance regions of the test procedures for the elementary hypotheses. The chief motivation behind this method, only partially explained in previous publications, is the following. The expression of a complex hypothesis as the intersection of a number of components is by no means unique, and, when we express it as the intersection of a particular class of components, we are looking for a test procedure that should have a large power against each of the associated deviations, may be at the cost of being relatively poor against other directions of deviation. At this point of the process we merely hope that this is precisely what would be achieved by the procedure suggested. For a wide class of specific problems in the setup of univariate and multivariate "normal" responses, it so turns out that this is, in fact, achieved by the test procedures suggested, in the sense that, in each case, the suggested procedure does better, for its particular purpose, than the one ordinarily used or recommended. At this point the further question arises. Even assuming that, for its particular purpose, the suggested procedure is better than the one ordinarily recommended, is it in any sense optimal (again for the particular purpose)?. To answer this question we have to define carefully the criterion (or sense) under which we are looking for optimality, and, in test procedures involving several parametric functions, that go with the more complex problems (as opposed to those like F, r, the two kinds of multiple correlation and Hotelling's T^2, etc., that involve only one parametric function each, and go with much simpler problems) any such criterion that might be laid down would seem to be far less "natural" and convincing than the corresponding one for simpler types of problems. By and large, the situation is this. In each case the procedure we offer can be shown to belong to a class that is good in a "natural" sense; and, among that class, the procedure can be shown to be optimal in a sense that is far less "natural" and convincing. Such results on the so-called "optimality"

as are known to date on these specific problems, and any light in this sense that we may be able to throw on the general method itself will be offered later. But in these complex situations, we must caution once more against any hasty attempt to set up an optimality criterion, then obtain an optimal procedure under that criterion, and finally stay happy with that procedure.

All this is for any given design. The power for all these procedures depend, in each case, upon α (the level of significance), p, v-1, n-v - b+1, the deviations, and the "error" dispersion matrix Σ_k which depends definitely upon the block size k (assumed, for simplicity to be the same, for all blocks), i.e., upon the design chosen. Holding n (the no. of experimental units), v (no. of treatments) and p (no. of variates) fixed, for any test procedure, a design with a smaller value of k will push up the contribution to the power from Σ_k and will pull down the contribution from the d.f. n-v-b+1, but the first effect (for a properly chosen design) will more than offset the second effect. Various useful designs have been discussed in [15].

Turning to the general design aspect of the total problem, we observe that, once we lay down the objective, it has been possible, at least for a wide class of problems, to suggest a test procedure that, under any design, in general, is good in the sense already explained (at this point no optimality is claimed). With this procedure at our disposal the further problem of choice of a good design, at least for these problems, would be governed by the possibility of (i) increasing contribution from Σ_k (a widely occurring empirical phenomenon well known to the design of experiments and ANOVA people, but apparently almost totally unknown to other groups of mathematical statisticians) and (ii) further increasing the discrimination in the preferred directions (already favored under the test procedure) through the structure of the design sought for, if necessary, by introducing some kind of asymmetry. The remarks made in the previous paragraph about "good" and "optimal" in

relation to a test procedure would equally apply to the choice of a design. As, for example, in the above problems, it is possible to choose a design that would be good in the sense of being better than the customary ones, but optimality is another matter. For a general theoretical treatment of optimality, in addition to the difficulties already mentioned in relation to the test procedure, we encounter the further difficulty about defining a sufficiently wide and meaningful class of "good" designs among which the one sought for is to be "optimal". However, in the sense explained in [15], we have a very restricted treatment of optimality. In other words, it is broadly indicated there how it is possible, on the basis of the criteria and techniques (i) and (ii) mentioned just now, to pick out an "optimal" design among a particular small class of designs each of which is "good". So far as the general treatment of optimality is concerned, the little that has been done, for whatever it is worth, will be discussed later. Finally, we would like to add the same note of caution as at the end of the last paragraph.

4. Concluding remarks. To illustrate certain principles for selecting a test procedure and a design, the foregoing section uses a two factor experiment in which one factor, consisting of a set of blocks, is unstructured, and the other factor, which may or may not be initially structured, eventually gets some kind of a structure imposed on it in that we are more interested in a certain set of contrasts than in other contrasts. Such preference, of course, becomes all the more natural and justified if the set of treatments comes from the setup of classical factorial experiments where all the factors are, in general, structured (except the set of blocks, when such a set is considered). For example, with a set of treatments at v levels, belonging to one (possibly) continuous factor, a very natural pattern of decomposition would be $\tau_i - \tau_{i+1}$ $(i = 1,2,\ldots,v-1)$ which gives, to a first approximation, the nature of the dependence of τ_i on 'i'. For factorial experiments of the

classical type (involving structured factors) the remarks of section 3 and the results of [15] are specially applicable and some further advances have also been made for such experiments that will not be reported here.

However, for such experiments there is another line of development with a different specific purpose, due to Gnanadesikan and Wilk : [4], in extension of a similar one due to Daniel, that is worth mention here. In [4] references are given to relevant previous literature. The following is an abstract of [4].

A procedure is presented for the generalization and extension, to multiresponse factorial experiments, of the technique of half-normal plotting for uni-response factorials.

Consider, for definiteness, two-level factorial experiments wherein, for each treatment combination, p responses are observed. For this multi-response situation, the analogue of the uni-response single degree of freedom contrasts is a vector of p elements, each element being a single degree of freedom contrast corresponding to one of the responses.

A positive semidefinite or definite quadratic form in the elements of each of these vectors is obtained (for example, the squared length of the vector). This is interpretable as a distance function in a meaningful space. The null distribution of the quadratic form is approximated as a gamma with two parameters. Under reasonable experimental assumptions, the quadratic forms are mutually independent. Using only the e smallest of the quadratic forms as a sample of the first e order statistics from a sample of size g (where $e \leq g \leq v-1$, the total number of contrasts) from a gamma distribution, maximum likelihood estimates are obtained for the parameters of the gamma distribution.

Using the estimates, a "gamma plot" is made of the ordered quadratic forms. Interpretations and uses of the plot are discussed, with examples.

REFERENCES

[1] Bhapkar, V. P., "Contributions to the statistical analysis of experiments with one or more responses (not necessarily normal)," Institute of Statistics, University of North Carolina, Mimeograph Series No. 229 (1959).

[2] Cox, D. R., Planning of Experiments, John Wiley and Sons, New York (1958).

[3] Fisher, R. A., The Design of Experiments, (Fifth edition), Hafner Publishing Company, Inc., New York (1949).

[4] Gnanadesikan, R. and Wilk, M. B., "Graphical analysis of multiresponse experimental data using ordered distances," Bell Tele.Labs.Series,1961.

[5] Kempthorne, O., The Design and Analysis of Experiments, John Wiley and Sons, New York (1956).

[6] Roy, J., "Step-down procedures in multivariate analysis," Ann. Math. Stat., Vol. 29 (1958), pp. 1177-1187.

[7] Roy, S. N., "A heuristic method of test construction and its use in multivariate analysis," Ann. Math. Stat., Vol. 24 (1953), pp. 220-238.

[8] Roy, S. N., Some Aspects of Multivariate Analysis, John Wiley and Sons, New York (1957).

[9] Roy, S. N. and Roy, J., "Analysis of variance with univariate or multivariate, fixed or mixed classical models," Institute of Statistics, University of North Carolina, Mimeograph Series No. 208 (1958).

[10] Roy, S. N. and Gnanadesikan, R., "Some contributions to ANOVA in one or more dimensions," Ann. Math. Stat., Vol. 30 (1959), pp. 318-340.

[11] Roy, S. N. and Bhapkar, V. P., "Some nonparametric analogues of 'normal' ANOVA and MANOVA and of studies in 'normal' association," Hotelling Dedicatory Volume, Stanford University Press (1960).

[12] Roy, S. N. and Cobb, Whitfield, "Mixed model variance analysis with normal error and possibly non-normal other random effects - Parts I and II," Ann. Math. Stat., Vol. 31 (1960), pp. 939-968.

[13] Roy, S. N. and Roy, J., "On testability in normal ANOVA and MANOVA with all 'fixed effects'," Calcutta Mathematical Society Jubilee Commemorative Volume (1958-59), issued 1961.

[14] Roy, S. N. and Mikhail, W., "On the monotonic character of the power functions of two multivariate tests," Institute of Statistics, University of North Carolina, Mimeograph Series No. 238 (1959).

[15] Roy, S. N. and Srivastava, J. N., "Inference on treatment effects and design of experiments in relation to such inference," Institute of Statistics, University of North Carolina, Mimeograph Series No. 274 (1961).