GRAPHICAL METHODS FOR INTERNAL COMPARISONS IN MULTIRESPONSE EXPERIMENTS

By M. B. Wilk and R. Gnanadesikan

Bell Telephone Laboratories

1. Introduction. One of the functions of a statistical analysis of data is to exhibit, in a summary and meaningful way, what the observations have to suggest. The objectives of statistical analysis are neither so narrow nor so formal as described and implied by some statistical theories of estimation and testing hypotheses. Certainly it is not usual, in the writers' experiences, to ask of a body of data only limited questions via a few narrow, albeit well-defined, statistical hypotheses or parameters. There is much value in informal statistical procedures, as excellently illustrated by many uses of scatter diagrams.

In particular, there is a long existent need for procedures to handle data involving multivariate responses in such a way that the resulting statistical summary and analysis (i) takes some account of the multivariate structure, and (ii) encourages insight into the experimental situation (as distinct from carrying out artificial and often pointless tests of hypotheses). The indefiniteness and complexity of objectives of statistical analysis of multiresponse data emphasize the need for general informal procedures which help to convey to the data-analyzer some of the information implicit in the data. (For additional discussion, see Tukey (1962, Section IV).)

The main intent here is to present in detail a proposed (cf. Wilk and Gnanadesikan (1961)) graphical statistical procedure for the informal assessment of certain types of multiresponse data.

Sections 2 and 3 review some procedures for uniresponse and multiresponse situations, respectively. Section 4 presents notations and a canonical description for experimental situations to which the method of analysis proposed in Section 6 may be applied. Section 5 discusses the use of Hotelling's T^2 in relation to the present objectives. The proposed method is presented in Section 6; certain of its features are discussed in Section 7. Section 8 gives some examples of uses of the method. General discussion and concluding remarks are given in Sections 9 and 10.

2. Background-univariate. The analysis of variance, in addition to being useful for testing hypotheses and estimating various components of variation, also serves as an over-all frame for considering organized data, independently of any formal statistical model.

The analysis of variance of univariate data may be viewed as a recoordinatization of the basic data with the objective of having the new coordinates reflect, in a meaningful way, a structure suggested by the conditions under which the data has been collected.

Received 20 May 1963; revised 11 November 1963.

The new coordinate system should depend on relevant considerations, but *not* on irrelevant ones such as a desire to isolate single degrees of freedom even when no interpretable basis for the individual contrasts exists. When there is no interpretable basis for further decomposition, the outcome of the analysis of variance transformation is generally in terms of squared distances related to collections of degrees of freedom.

A usual formulation of the problems of analysis of variance involves assumed linear models and formal tests of hypotheses and estimation. [See for example, Graybill (1961), Scheffé (1959).] This process usually depends crucially on the selection of a valid error term.

Even with respect to apparently well-defined specific questions, differing viewpoints exist on the interpretations of the procedure employed. Thus some statisticians regard an F test in the framework of the Neyman-Pearson (1933) theory as a means of literally accepting or rejecting precise hypotheses; others regard it as a means of objectively evaluating the statistical significance of the evidence against certain "natural" null hypotheses.

Clearly, procedures which are limited by formal interpretation are not sufficient for the analysis of data. Informal appreciation is accomplished through various procedures including the summary analysis of variance table, study of residuals [Terry (1955), Anscombe & Tukey (1963)], tests for nonadditivity [Tukey (1949a)], tables and plots of means and standard errors, and the techniques of half-normal plotting [Daniel (1959)] and of gamma plotting [Wilk et al. (1962a)].

Analysis of variance problems usually involve many questions to be asked of the same body of data. In responding to this in the framework of formal probabilistic assessments many alternative and controversial systems have been advocated in what is known as the problem of multiple comparisons. [See, for example, Tukey (1953), Scheffé (1959).] These approaches involve simultaneous parallel comparisons with some "external" standard (for example, a preselected estimate of error).

In the opinion of the present authors, a useful objective of the analysis of variance is the joint relative assessment of comparable quantities—as for example a collection of contrasts in a 2^n experiment. Procedures which involve simultaneous comparisons among comparable quantities with the aid of a statistical measure or standard for facilitating interpretation, may be called *internal comparison* procedures. The statistical measure or standard may in part be externally generated but should always in part be internally generated.

Some examples of internal comparison procedures are: (i) the scatter diagram. Here the standard of continuity may be external while shape is internal; (ii) the half-normal plot [Daniel (1959)]. Here normality is external; particular straight line configuration is internal; (iii) spectral analysis [Blackman and Tukey (1958)]; peculiarities, such as peaks, are indicated comparatively; (iv) probability plotting and its various uses [Blöm (1958)].

3. Background-multivariate. Comparisons and assessments of treatment factors and their relationships, when the response is multivariate is intrinsically complicated in that there is, in general, no unique linear ordering for vectors. One approach is to associate a measure of size with each vector. [See, for example, Hotelling (1931), Mahalanobis (1936), and Fisher (1938).] Another approach is to attempt dimensionality reduction by a data-motivated transformation. [See, for example, Hotelling (1933) and Shepard (1962a, 1962b).]

The definition of treatment effects and interactions is multidimensional when the response is multivariate. For formal statistical assessments of the multidimensional treatment effects, various analogues of univariate procedures have been proposed.

Most attention has been directed to tests of significance. Especially in the multivariate case, such tests, for a null hypothesis of no effects against the completely general alternate hypothesis, have important limitations. Also other complexities arise. Thus, the requirement of invariance leads only to a class of tests that depend on the roots of certain determinantal equations [Wilks (1932), Lawley (1938), Roy (1939), Hotelling (1947), Tukey (1949b)]. Choice amongst these is currently based largely on intuition.

Confidence estimation procedures, of varying degrees of appropriateness, including a multivariate multiple comparisons procedure, have been proposed [Roy and Bose (1953), Roy and Gnanadesikan (1957, 1958), Anderson (1962)].

Roy (1958) has proposed a class of "step-down" procedures for confidence estimation and tests of significance. These involve combined consideration of the marginal behavior of one response and a hierarchical sequence of conditional behaviors of the other responses. The conclusions drawn depend upon the chosen order of the responses.

A two-sample test of significance for the case when the number of responses exceeds the number of degrees of freedom available for the estimation of error, has been proposed by Dempster (1958).

It is fairly common in analyzing multiresponse data formally to ignore the correlation structure between the responses, to analyze each response separately and then, intuitively, to integrate the conclusions arrived at from the several separate analyses. A variant of this approach is to begin with an initial principal components analysis and follow with uniresponse analyses of variance on each of the principal components. While this technique may lead to useful results in some problems, difficulties in interpretation often arise. [See, for example, Finney (1956), Gabbe and Wilk (1960), Roy et al. (1962), Wilk et al. (1962c), Yates and Healy (1951).]

Thus some formal analogues of univariate analysis of variance procedures, and some new problems, have been studied for the multiresponse case. A comprehensive review of multivariate methods has been given by Bartlett (1947). However, even such formal questions as the use of the characteristic vectors along with the characteristic roots, [cf. Smith et al. (1962)] need further investigation.

Unlike the univariate situation, in the multivariate case many of the methods developed to answer specific formal problems (often quite artificial) have not proved generally useful in providing interpretable summaries of the data. The present authors feel there is a need for additional techniques which may augment informal insight into the structure of multivariate data.

4. The present problem; experimental situation and notation. The concern of the present paper is to present a graphical internal comparisons procedure for the multiresponse single-degree-of-freedom case. These procedures constitute a generalization and extension of the technique of half-normal plotting which has been advocated by Daniel (1959) for analyzing certain single response data.

A prototype of the experimental situations for which the methods described below were developed is a 2^n factorial experiment in which each treatment combination leads to a multivariate response.

More generally, the methods are directly applicable and appropriate where the experimental design has such structure and objectives that, if consideration were limited to a single response, a meaningful analysis of the observations would involve an orthogonal transformation associated with an analysis of variance based on orthogonal single-degree-of-freedom contrasts.

For the sake of simplicity and definiteness, the subsequent developments will be presented in terms of a multiresponse 2^n factorial experiment. Generalization to fractionated and confounded 2^n factorials, and to other appropriate circumstances, is immediate.

A 2^n factorial experiment involves the testing of all possible treatment combinations of n factors, each at two levels. In many such experiments, when each of the 2^n distinct treatment combinations is applied to an experimental unit several responses are generated, such as (purity, yield, color, density, etc.). It will often be true that these observations will be subject to a joint statistical variability—will be statistically correlated—as well as, perhaps, their "true" values being physically interdependent. The statistical dependence of these jointly observed responses makes it desirable that the analysis of the data should take that dependence into account.

In contrast to the statistical correlations among the several responses observed for a single treatment combination on a given experimental unit, it will often be reasonable to assume that the responses, univariate or multivariate, from separate applications of treatment combinations to different experimental units will be statistically independent.

Thus suppose the treatment combinations to be numbered from 1 through $N = 2^n$, and suppose that p responses are observed on each treated experimental unit. Let the p-variate response to the ith treatment combination be denoted by

$$\mathbf{y}_i' = (y_{i1}, y_{i2}, \cdots, y_{ip}), \qquad i = 1, 2, \cdots, N.$$

The outcome of the entire experiment may be written as an $N \times p$ matrix $\mathbf{Y} = (y_{ij})$, $i = 1, 2, \cdots, N$; $j = 1, 2, \cdots, p$, whose rows are the $\mathbf{y}_i'$ and whose columns may be denoted by $\mathbf{Y}_j$, $j = 1, 2, \cdots, p$.

It will often be experimentally reasonable to treat the p-dimensional response vectors, $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$, as though they were observations on mutually independent random vectors, with the assumption, usually not critical, that all of them have the same covariance matrix. Hence,

$$E[(\mathbf{y}_i - E\mathbf{y}_i)(\mathbf{y}_j - E\mathbf{y}_j)'] = \delta_{ij}\mathbf{\Sigma}(p \times p), \qquad \text{for } i, j = 1, 2, \cdots, N,$$

where δ_{ij} is the Kronecker delta.

When analyzing an individual response in such an experiment, it is usual and meaningful to transform, orthogonally, the N observations for each response to a set of $(N - 1)$ contrasts, usually called main effects and interactions, plus one number expressing the general mean. Thus, for the jth response, the main effects and interactions, in some desired order, and the mean, would be given by,

$$(\mathbf{X}_j', m_j) = \mathbf{Y}_j'\mathbf{R} = (x_{1j}, x_{2j}, \cdots, x_{N-1,j}, m_j), \qquad j = 1, 2, \cdots, p$$

where $\mathbf{R}$ is an appropriate $N \times N$ orthogonal matrix.

If the rows of the matrix $\mathbf{Y}$ each have the same covariance matrix $\mathbf{\Sigma}$ and are mutually uncorrelated, then because $\mathbf{R}$ is orthogonal, the same properties will hold for the rows of the matrix $\mathbf{X}$ given by the $(N - 1) \times p$ matrix $\mathbf{X} = (x_{ij})$ whose rows are denoted by $\mathbf{x}_i'$ and whose columns are the $\mathbf{X}_j$.

For each response, one of the accomplishments of the transformation of the observations to main effects and interactions is that when the treatment combinations do not, in fact, have differential effects then the main effects and interactions will, aside from statistical fluctuations, be measures of zero. Hence, under such null assumptions, which often serve admirably as a basis for their statistical contradiction,

$$Ex_{ij} = 0, \qquad i = 1, 2, \cdots, N - 1; j = 1, 2, \cdots, p.$$

Furthermore, because each x_{ij} is obtained as a linear combination of random variables $y_{1j}, y_{2j}, \cdots, y_{Nj}$, assumed independent, it is a reasonably robust assumption that the contrast vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N-1}$, may be treated as a random sample from multivariate normal distributions, each with the same covariance matrix $\mathbf{\Sigma}$, and, under the null assumption, with common mean vector $\mathbf{0}$. It may be noted that, even if the response vectors $\mathbf{y}_i$, $i = 1, 2, \cdots, N$, do not have the same covariance matrix, the covariance matrices of the contrast vectors $\mathbf{x}_i$, $i = 1, 2, \cdots, N - 1$, will all be equal, so long as the $\mathbf{y}_i$ are uncorrelated.

5. Hotelling's T^2 test. A significance test for the multivariate null hypothesis that a particular main effect or interaction has no real effect, i.e. that $E(\mathbf{x}_u) = \mathbf{0}$ for a specific u, is provided by Hotelling's T^2 statistic [Hotelling (1931)], $T_u^2 = \mathbf{x}_u'\mathbf{S}^{-1}\mathbf{x}_u$, where $\mathbf{S}$ is an unbiased estimate of $\mathbf{\Sigma}$, statistically independent of $\mathbf{x}_u$, and has a Wishart distribution.

Methods for obtaining the estimate $\mathbf{S}$ deserve some discussion. In 2^n multiresponse experiments, it may be possible to specify, from prior physical knowledge or intuition, certain effects as being "negligible." If so, then such "error" contrast vectors, can be combined to provide the estimate $\mathbf{S}$. (The matrix $\mathbf{S}$

will, in general, be non-singular if the number of linearly independent vectors used in obtaining it exceeds p, the number of responses.)

Alternatively, if some treatment combinations are observed on more than one experimental unit then these replications "within cells" may be used to generate an S.

By whatever legitimate process the estimate S is obtained however, it should be noted that:

(i) The unconditional (on S) distribution of T_u^2 does not involve Σ.

(ii) Conditional upon S, the distribution of T_u^2 involves the unknown covariance matrix Σ.

(iii) Quadratic forms T_u^2, for various values of u, all computed using the same matrix S, are mutually statistically *dependent* unconditional on S.

(iv) For fixed S, but unconditionally varying $\mathbf{x}_u$, the various T_u^2 are statistically independent and, under null assumptions, will behave like a random sample from a common distribution which will indeed involve Σ as well as S. This is discussed further in the next section.

6. The proposed method. An important contribution of the technique of half-normal plotting, [Daniel (1959)], is that it provides a graphical, interpretable summary, which facilitates simultaneous comparisons among statistically comparable quantities—in that case, the single-degree-of-freedom contrasts which estimate the effects in a 2^n uniresponse experiment. This objective is not attained by separately testing the significance of each contrast by t-tests nor by the usual multiple comparisons procedures.

Similarly, for the multiresponse 2^n situation, the use of separate Hotelling's T^2 tests on the contrast vectors, or of certain multivariate multiple comparisons methods, will not supply a basis for the simultaneous intercomparison of the contrast vectors. The present section describes a statistical procedure for obtaining graphical internal comparisons appropriate to the analysis of 2^n multiresponse experiments.

The outcome of the experiment may be represented as N points in the p-dimensional response space, one point corresponding to each treatment combination. The usual analysis into main effects and interactions defines, on this set of points, $(N - 1)$ distinct partitions into equal halves. For each effect, the centroid of each of the two sets (halves) of points, defined by the corresponding partition, may be found. The uniresponse effects correspond to the distances between the projections of these two centroids onto the individual response axes. The squared Euclidean distance between the centroids in p-dimensional space is then the sum of squares of the uniresponse effects. Figure 1 gives an illustration for the A main effect of a 2^3 experiment with $p = 2$ responses per experimental unit.

The $(N - 1)$ contrast vectors may, more generally, be visualized as $(N - 1)$ points in p dimensions. By the definition of a metric in this space, a length, a distance from the origin, may be associated with each of these contrast vectors.

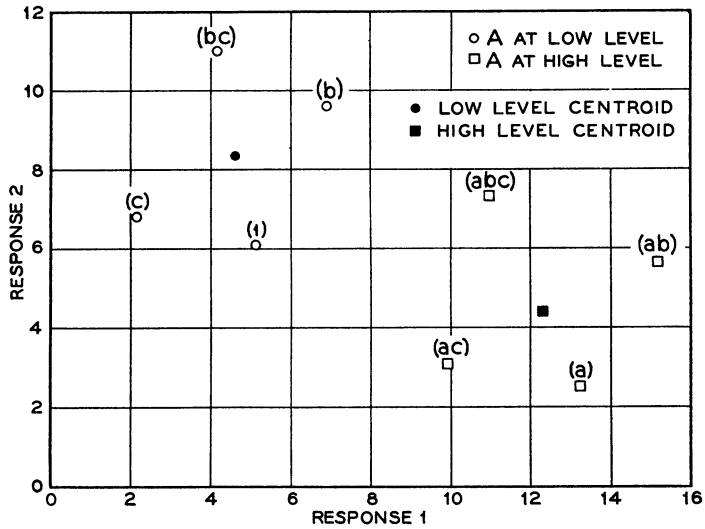


FIG. 1. Geometric representation corresponding to A main effects in a 2^3 experiment with 2 responses.

The metrics considered here are the square roots of non-negative quadratic forms in the elements of the contrast vectors. For a chosen compounding matrix $\mathbf{A}$, the squared distance associated with the contrast vector $\mathbf{x}$ is $d = \mathbf{x}'\mathbf{A}\mathbf{x}$.

There is a direct relation between the two preceding geometric representations. The measure of size associated with a contrast vector is interpretable as a squared distance, for an appropriate metric, between the centroids of the sets of vector observations defined by the contrast. In future reference, the term "distance" will be used in place of "squared distance". For instance, using $\mathbf{A} = \mathbf{I}$, the identity matrix, corresponds to the use of the squared Euclidean distance between the centroids.

Some other examples of appropriate compounding matrices are: (1) A diagonal matrix of reciprocals of variances (specified or estimated) of the p responses. (2) The inverse of the covariance matrix (specified or estimated) of the original responses. These may be particularly appropriate when the responses are measured on scales yielding widely differing variances. Other kinds of compounding matrices of either general or particular applicability can be useful.

Whatever the compounding matrix used, it is the general philosophy of the present approach that subsequent inferences be "conditional" on this choice.

For a selected compounding matrix $\mathbf{A}$, each of the distances $d_i = \mathbf{x}_i'\mathbf{A}\mathbf{x}_i$, $(i = 1, 2, \cdots, N - 1)$, will be distributed as a linear combination of r mutually independent single-degree-of-freedom chi-squares (central under the null assumptions of no real effects), where the coefficients are the positive characteristic roots of $\mathbf{A}\boldsymbol{\Sigma}$, and $r \leqq p$ is the rank of $\mathbf{A}$. Clearly, when $\mathbf{A} = k\boldsymbol{\Sigma}^{-1}$, each distance has a $k\chi^2(p)$ distribution.

In application, $\mathbf{\Sigma}$ is unknown while $\mathbf{A}$ is chosen by the user. Hence the distribution of each d_i will depend on the unknown characteristic roots of the matrix $\mathbf{A\Sigma}$.

In the method described below, the inferential process is "conditional" upon the particular matrix $\mathbf{A}$ employed in the distance functions $\{d_i\}$ and depends upon the *estimation* of the distribution of d_i. Hence the description of the method does not depend on how $\mathbf{A}$ is selected or whether it originates from quantities which are, in a larger reference set, random variables. (See Section 9(c) for further discussion.)

Thus—whether $\mathbf{A} = \mathbf{I}$, or $\mathbf{A}$ is selected to reflect "economic" reasons, or $\mathbf{A}^{-1}$ is an independent unbiased estimate of $\mathbf{\Sigma}$ (in a larger reference set), or $\mathbf{A}^{-1}$ is any other type of estimate of $\mathbf{\Sigma}$—considering the specific $\mathbf{A}$ as fixed, the distribution of d_i will be as indicated above. For the experimental situation described above, the joint *null* distribution of the quantities $d_1, d_2, \cdots, d_{N-1}$, for fixed $\mathbf{A}$, will be that of a random sample of size $N - 1$ from such a distribution.

It has been suggested, among others, by Satterthwaite (1941), Patnaik (1949), and Box (1954), that such a distribution may be satisfactorily approximated by $\mu\chi^2(\nu)$ for suitably chosen μ and ν.

One procedure which depends upon knowing the characteristic roots of $\mathbf{A\Sigma}$ (not available in the present application), is to equate the means and variances of the two distributions.

Another procedure for determining ν, which is also not applicable in the present context for reasons implicit in the discussion below, has been proposed by Dempster (1958).

Under null assumptions of no real treatment effects, the $(N - 1)$ distances $d_1, d_2, \cdots, d_{N-1}$ may, as a reasonable approximation, be expected to behave like a random sample from a gamma distribution with origin parameter zero, unknown scale parameter λ and unknown shape parameter η.

If a "proper" estimate (discussed below) of η were available then a meaningful statistical summary would be obtained by plotting the $N - 1$ *ordered* distances versus the appropriate quantiles of the standard ($\lambda = 1$) gamma distribution. A procedure for carrying out this gamma probability plotting has been described by Wilk et al. (1962a).

Of course, just as the half-normal plot can be applied to subsets of contrasts, the entire procedure may be applied to reasonably chosen subsets of the $(N - 1)$ distances.

If the null assumption is correct, and η properly chosen, then the resulting plot should tend to appear as a straight line, passing through the origin with slope $1/\lambda$. Note that the procedure depends only on having an estimate of η; the scaling factor λ influences only the slope and not the collinearity of the points.

If the null assumption is not borne out by the data, then the largest d_i values will be "too large" and will appear as major departures or curvature away from a straight line configuration.

Since η is unknown its value must be estimated. To minimize the influence

of *possible real* effects on this estimation, it is based on an order statistics formulation. The crux of the issue is that some of the distances may *not* obey the null assumptions. The basic object of the procedure is to allow them to (statistically) exhibit themselves, relative to the remaining quantities.

Under nonnull conditions the distances will involve noncentral chi-squared variates which can *still* be approximated by a gamma distribution. [Cf. Patnaik (1949).] Hence, the inclusion of *all* the distances in the process of estimation of η would tend to generate a value for η which would obscure the fact that some of the distances are statistically "too large." However, it will usually be reasonable to assume that a number of the smallest distances (order statistics) do essentially satisfy null assumptions.

With this in mind, the following procedure is adopted:

L contrast vectors, which one is interested in studying comparatively, are chosen ($L \leqq N - 1$) and, for a selected compounding matrix, the associated distances calculated. Using judgment, a number $K(\leqq L)$ is assigned as the number of contrast vectors which may well not reflect systematic effects. The $M(\leqq K)$ smallest distances, which are thereby even less likely to reflect systematic effects, are then considered as the M smallest of a random sample of size K from a gamma distribution. The estimation of η and λ is based upon this statistical formulation. One possible approach is via maximum likelihood. The maximum likelihood estimates of η and λ depend only on the Mth largest distance and on the arithmetic and geometric means of the M distances. The problem of the maximum likelihood estimation of η and λ from the first M order statistics in a random sample of size K from a gamma distribution has been discussed and tables provided in Wilk et al. (1962b).

Next, using the estimate $\hat{\eta}$ of η, a plot is made of the L ordered distances against the quantiles of the standard gamma distribution ($\lambda = 1$) with $\eta = \hat{\eta}$. Under the null assumption of no real effects, one would expect to obtain a straight line pattern with intercept 0 and slope $1/\lambda$. Distances corresponding to real or systematic effects will tend to appear as "too large" deviations from the straight line pattern.

7. Discussion of certain features of the method.

(a) *Choice of the effects to be compared.* The user will, in most circumstances, be well advised to "censor" the set of effects whose associated distances are to be plotted. This idea is much the same as the process of "nominating contrasts" in the analysis of variance, proposed by Pearce (1953) and advocated by Daniel (1959), following Tukey, in connection with half-normal plotting. Similarly, in most factorial experiments, it will often be revealing to partition the effects according to criteria such as order of the interaction or involvement of specific treatment factors. The choice of L may be based, at least in part, on such considerations.

(b) *Choice of K and M for estimating η.* The method for the estimation of η deliberately leaves the choice of M and K indefinite. The overall procedure is

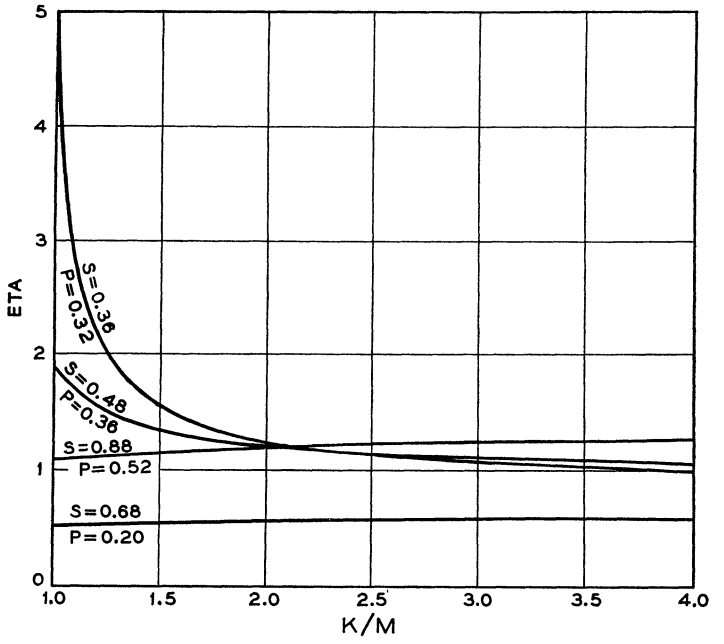


FIG. 2. Dependence of $\hat{\eta}$ on K/M for various P and S values.

advocated specifically as an informal method to aid the understanding of the experimental results. As such, it cannot and should not attempt to be independent of prior as well as posterior (after seeing the data) judgment.

In this estimation procedure, two sources of bias are distinguishable, namely, incorrect choice of K, and improper inclusion among the M smallest distances of distances involving real effects. Of these, bias from the latter source is more likely to affect the estimate of η so seriously as to distort the statistical implications of the configuration of the plotted distances.

The estimation of the parameters λ and η, for fixed M, is rather insensitive to the value of K provided M is not too close to K. This is illustrated in Figure 2, (based on Wilk et al. (1962b)), for various values S and P, where S and P are respectively the arithmetic and geometric means of the M smallest distances each divided by the Mth distance. For many 2^n factorial experiments, a choice of M such that K/M is greater than $\frac{3}{2}$ will be quite safe. In most situations, the loss of efficiency, in estimation of η, due to choosing M small will have little effect on the graphical internal comparisons.

(c) *Choice of the compounding matrix.* It is recommended that, usually, several different compounding matrices should be employed. Clearly a nondegenerate multidimensional situation cannot be entirely described in a single one-dimensional representation. Each distinct distance function plot gives a different insight into the factorial structuring of the data. The ill-defined problem of how and

why to choose various compounding matrices, discussed in Wilk et al. (1962c), is under continuing study.

(d) *The need for replotting.* An integral part of the use of this method, as it is of half-normal plotting, is the realization that several phases of replotting may be desirable. Thus, if an initial plot indicates statistical evidence of some real effects, then the existence of these will affect the entire configuration. Replotting after omission of these "too large" distances may enable one to see other subtle peculiarities.

In summary then, it will usually be desirable and appropriate to reanalyze the data with different choices of L, K and M, as well as according to several choices of the compounding matrix.

The proposed method is expected to be useful in the assessment of real effects and, as in the case of the half-normal plotting procedure, in the identification of other causes for departures from the straight line pattern, as for example, bad nonnormality and the operation of more than one error covariance matrix.

8. Some examples. Some aspects of the method are illustrated using data from three examples. No detailed discussion of the examples *per se* is intended.

Example 1 concerns experimental data reported by Daniel and Ribblet (1954) for a 2^{8-3} experiment involving $p = 2$ responses, called activity and selectivity, for each experimental unit. Example 2 deals with computer generated "random multivariate normal deviates." Sixty independent 5-dimensional vectors were generated with mean the null vector and covariance matrix,

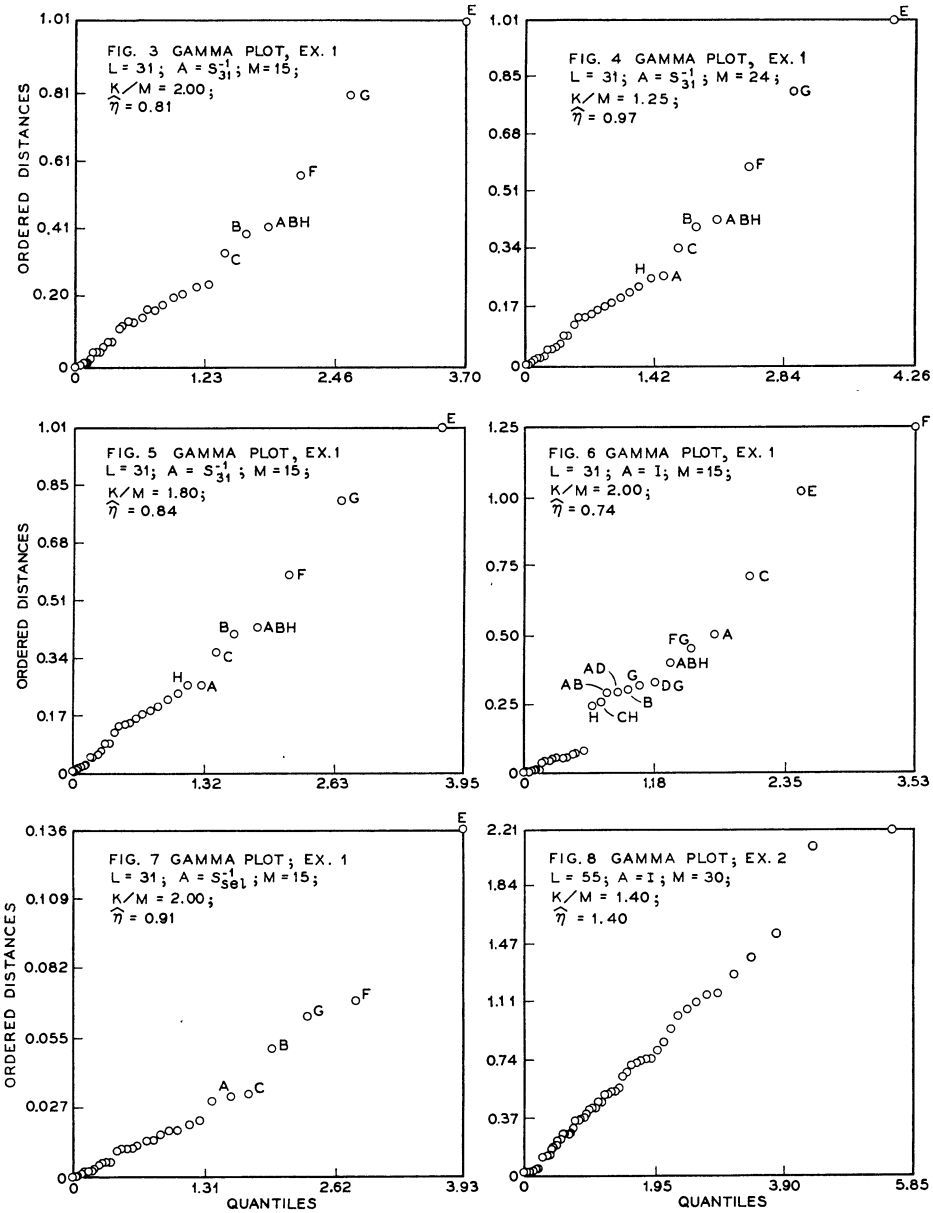
$$\Sigma = \begin{pmatrix} 1 & & & & \\ 2 & 5 & & & \\ -3 & -7 & 11 & & \\ 4 & 10 & -16 & 25 & \\ 2 & 5 & -8 & 9 & 16 \end{pmatrix}.$$

To a random selection of ten of these were added the following noncentrality vectors: Three of (3, 7, 10, 12, 11), three of (5, 5, 5, 5, 5), three of (7, 2, 0, 5, 4), and one of (5, 8, 15, 20, 18). Example 3 concerns experimental data from a video-telephone study (Gabbe and Wilk (1960)) organized as a 2^{9-1} in a split-plot design with $p = 8$ responses per experimental unit. Each of these responses involved a subjective assessment of picture quality.

(a) *Effect of choice of M for fixed K.* Figures 3 and 4 are plots of the $L = 31$ distances associated with the contrast vectors of Example 1. These distances are based on a compounding matrix which is the inverse of a sums-of-products matrix involving all 31 contrast vectors. For Figure 3, the estimate of η was based on $M = 15$, for Figure 4 it was derived from $M = 24$. In each case the value $K = 30$ was used.

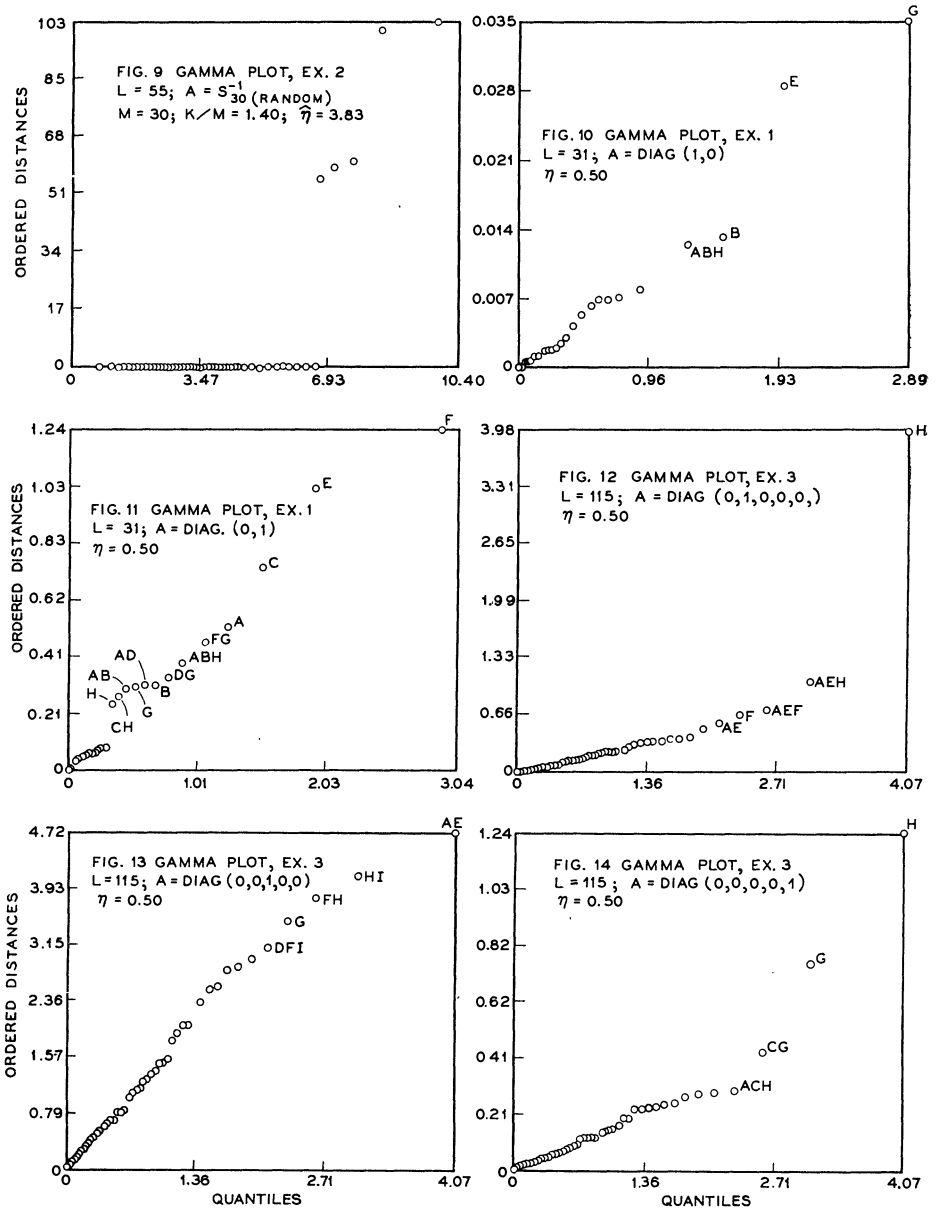
It will be seen that the effect of changing M for fixed K is negligible so far as patterns evident in the statistical configuration of the distances are concerned.

(b) *Effect of choice of K for fixed M.* Figures 3 and 5, also based on Example 1,



show the plots for choices of $K = 30$ and $K = 27$, respectively, for common value of $M = 15$. The two configurations are very similar.

(c) *Effect of choice of compounding matrix.* Figures 6 and 7 give internal comparisons plots for Example 1, for two choices of the compounding matrix, namely, the identity matrix $\mathbf{I}$ and $\mathbf{S}_{sel}^{-1}$, the inverse of a sums-of-products matrix based on the 23 two and three factor interaction contrast vectors. Figure 6 shows a



split into two groups of 18 (smaller) and 13 (larger) distances. Among the 13 (larger) distances, 7 out of 8 main effects are included.

Figure 7, on the other hand, is much smoother and leads to segregation of the main effect E as the only possibly real effect.

Figures 8 and 9 are based on Example 2, the two compounding matrices being the identity matrix $\mathbf{I}$ and the inverse, $\mathbf{S}^{-1}$, of a sums-of-products matrix based

on a random selection of 30 out of the 50 "central" contrast vectors. Each figure is a plot of the smallest 55 distances. Figure 8, based on **I**, does not lead to detection of the 5 distances associated with the 5 nonnull contrast vectors known to be present among the 55 plotted points. On the other hand, Figure 9, based on $\mathbf{S}^{-1}$, clearly delineates the 5 noncentral points.

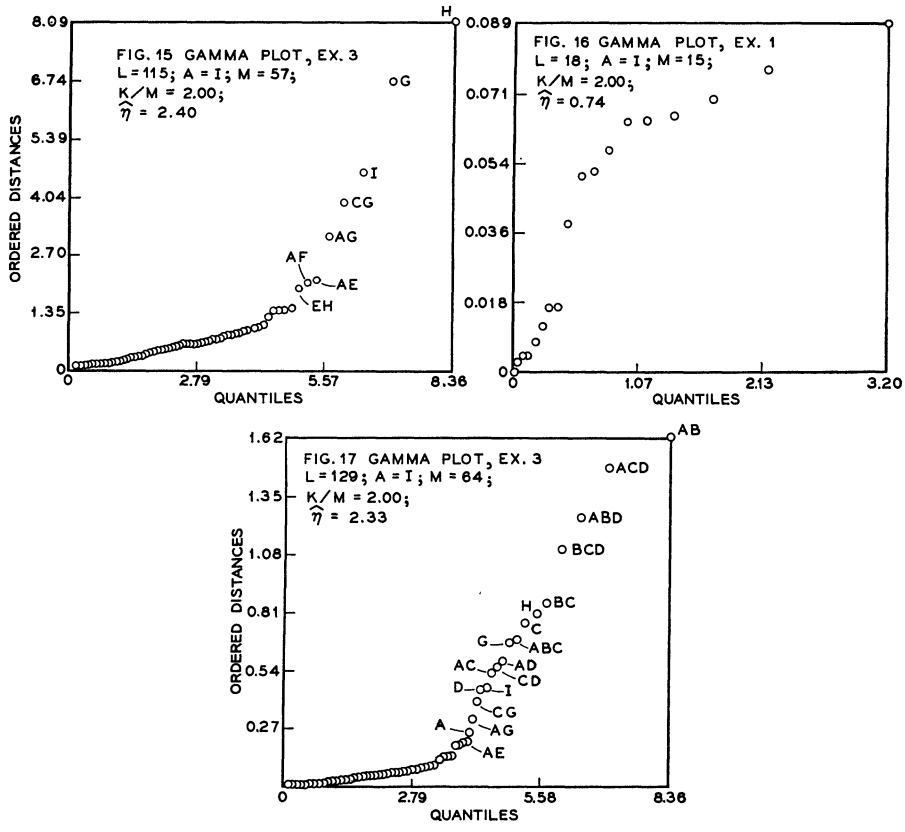
(d) *Relation of analyses of various subsets of the responses.* The present method is not meant to displace preliminary analyses of the separate responses through half-normal plotting (cf. Daniel (1959)). A configuration entirely equivalent to the half-normal plot may be obtained as a special case of the method proposed here. Thus, if the compounding matrix is taken to be diagonal of rank 1 where the nonzero element corresponds to the response under analysis, and if the value of η is taken as $\frac{1}{2}$, then the gamma plot so obtained will be equivalent to a $\chi^2(1)$ plot of the squared contrasts for that response. (In some circumstances, even with uniresponse data, the use of an estimated η, instead of $\eta = \frac{1}{2}$, may be an interesting supplement.)

Two such plots, for Example 1, are shown in Figures 10 and 11, corresponding to the two separate responses of the experiment. These plots are rather ragged, exhibiting pronounced curvatures even near the origin. From Figure 10, for the first response, one might reasonably suspect two real effects, (main effects E and G). In Figure 11, for the second response, there appears a split configuration similar to that of Figure 6 (based on **I** as compounding matrix). A joint consideration of Figures 6, 7, 10 and 11 suggests the possibility that among the individual observations for the second response there might have been two which were overly "large" in absolute value. A more definite assessment of this indication could not be made because the original data are not available.

Figures 12, 13, 14 and 15, based on Example 3, illustrate how multiresponse analysis can importantly augment the separate uniresponse analyses. Figures 12, 13 and 14 are typical of the eight $\chi^2(1)$ plots of the squared contrasts for the separate responses of the experiment. Some of these plots, e.g. Figures 12 and 14, suggest the existence of one or two real effects. Others, as Figure 13, do not indicate any real effects. On the other hand, Figure 15 based on a multiresponse analysis, using **I** as compounding matrix, indicates seven or eight possible real effects.

(e) *Effect of replotting.* Graphical assessment will often benefit from replotting, after omitting "too large" points. Some care in this is desirable but appreciation is developed with experience. Figure 16 shows a replot from Example 1 of the smallest 18 points in Figure 6. While Figure 6 conveys the impression that the lower 18 points form a homogeneous "error" set, Figure 16 is very ragged, shows a further faintly suspicious separation, and suggests additional possible peculiarities.

(f) *Possible other uses.* An example of an auxiliary value of the present method, similar to that of half-normal plotting, is given by Figure 17, based on Example 3. In this plot are shown 129 distances corresponding to the main effects and two and three factor interactions of the nine factors. This was a split-plot experiment



and 14 of the 129 contrast vectors had a whole plot error covariance matrix while the remainder had a different error covariance matrix. This distinction is evidenced in Figure 17 with its configuration suggestive of two intersecting straight lines. Among the points of Figure 17 which appear to lie on the line of steeper slope are all 14 of the distances corresponding to contrast vectors (involving factors A, B, C and D) with the whole plot error covariance matrix.

9. General discussion. (a) The proposed method is meant to provide a point of view in statistical analysis additional and analogous to that of half-normal plotting for the various responses separately or for various special linear combinations of these. The multiresponse method can importantly augment the separate marginal analyses for at least two reasons: (i) permitting the accumulation, in the distance functions, of smallish real effects in the individual responses; (ii) allowing the possibility that the statistical correlations within the observation vectors will exert a stabilizing influence in the analysis of appropriately chosen measures of size.

The multiresponse distances often lead to smoother and more stable statistical

configurations than the uniresponse analyses. This also may be due to a stabilizing effect from the intercorrelations.

(b) Aside from the need for distributional approximation, the distinctive departures in the multiresponse case from the uniresponse case are the freedom of choice of the compounding matrix and the need to estimate η, for making the plot.

(c) The intent of the method is to provide a system of internal comparisons and, as such, any aspects which are common to all elements being compared are properly regarded as part of the common fixed background for the comparison. Thus, for example, in the internal comparison of uniresponse contrasts, which might all be divided by an estimate, s, of the standard deviation, an appropriate probability plot would be half-normal and *not* "half-Student's t."

The configuration of the plot and the indications therefrom depend, of course, on the particular compounding matrix employed in the distances. The compounding matrix is, however, common to all of the contrast vectors which are being internally compared. And, the actual distances are employed in the estimation of the appropriate gamma distribution. Hence, the validity of the method does not depend on how the compounding matrix is selected or whether it originates from quantities which are, in a larger reference set, random variables.

(d) The use of a compounding matrix having rank 1 corresponds to the analysis of a particular linear function of the different responses. As indicated in Section 8, separate analyses of the individual responses is a special case of such a choice. The use of a compounding matrix having rank greater than 1 does not correspond to the analysis of any single linear function.

(e) It should be noted that the basic procedure is not dependent upon having an estimate of λ, just as half-normal plotting does not depend upon the knowledge of the error standard deviation.

(f) The present method depends upon estimating the value of η, for the evaluating distribution, from the data. The configurations obtained from the present statistical procedure appear to be relatively stable for reasonable variations in the value of η used in making the plot.

(g) In use of probability plotting it is well to keep in mind that the statistical stability of the configuration is very uncertain when the number of points involved is small. [Cf. Wilk et al. (1962a)].

(h) It is the experience of the present authors that such probability plots provide a useful summary of indications and a stimulus for insight. The attempt always to interpret these plotting procedures as formal significance testing procedures would be misguided.

(i) The present method, as also the half-normal plotting procedure, has something of the flavor of an informal "multiple comparisons" process: whether a quantity looks "large" depends on how many items are in the set. But the method differs importantly in the following ways from formal multiple comparisons procedures:

(1) The basic comparisons are internal among the actual distances rather than

being simultaneous assessments of differences between subgroups against some *external* measure (as for example, in the comparison of treatment differences with a preselected estimate of error variance).

(2) Repeated partitioning of the contrast vectors and of the responses and replotting are suggested and stimulated by the method.

(3) A motivation for methods of internal comparisons is to seek aid in gaining insight into the structure of the data. Such methods do not seem to belong with the rigid interpretations of the theory of testing hypotheses.

(j) Even where one is interested in giving formal answers to specific questions, informal procedures, such as the one proposed heretofore, may be useful tools. For instance, if tests or confidence regions depending on an estimate of the error covariance matrix are desired, then it may be wise to make a preliminary assessment (say by gamma plots of distances) to decide which effects might be sensibly employed in the estimation.

(k) There are various uses of the present graphical method in addition to the aspect of detection of real treatment effects. One such use has been illustrated in Section 8 in detecting the existence of two error covariance structures. Another application has been made by Laue (1961) in multivariate screening of electronic components. Other applications such as the detection of bad multivariate non-normality are under continuing study.

10. Concluding remarks. In the analysis of variance framework, several areas of need for graphical internal comparisons procedures may be distinguished. These are summarized in Table 1.

Cell I has been dealt with by Daniel (1959); Cell II has been discussed by Wilk et al. (1962); Cell III is under preliminary study by Wilk and Gnanadesikan (1963); Cell IV has been the major concern of the present paper; Cells V and VI remain for future detailed investigation.

11. Acknowledgments. It is a pleasure to acknowledge the assistance of Miss Elizabeth Lauh and Miss Marilyn J. Huyett in the development of examples, and comments of D. R. Cox, C. Daniel, M. J. R. Healy, C. L. Mallows, R. S. Pinkham and J. W. Tukey which were of help in revising an earlier draft.

TABLE 1

Response structure

Decomposition of factor structure	Uniresponse	Multiresponse
All 1 d.f. contrasts	(I) Half-normal $\sim$ Gamma ($\eta = 1/2$) $\sim \chi^2(1)$	(IV) Gamma (estimated η)
All ν (>1) d.f. groupings	(II) Gamma ($\eta = \nu/2$) $\sim \chi^2(\nu)$	(V) ?
Mixed d.f. groupings	(III) Under Study	(VI) ??

REFERENCES

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

ANDERSON, T. W. (1962). Some confidence bounds for determinantal roots. Abstract. *Ann. Math. Statist.* **33** 297.

ANSCOMBE, F. J. and TUKEY, J. W. (1963). The examination and analysis of residuals. *Technometrics* **5** 141–160.

BARTLETT, M. S. (1947). Multivariate analysis. *J. Roy. Statist. Soc. Suppl.* **9** 176–197.

BLACKMAN, R. B. and TUKEY, J. W. (1958). *The Measurement of Power Spectra.* Dover, New York.

BLÖM, GUNNAR (1958). *Statistical Estimates and Transformed Beta Variables.* Wiley, New York.

BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance–I. *Ann. Math. Statist.* **25** 290–302.

DANIEL, CUTHBERT (1959). Use of half-normal plots in interpreting factorial two level experiments. *Technometrics* **1** 311–341.

DANIEL, CUTHBERT and RIBBLET, EARL W. (1954). A multifactor experiment. *Ind. Eng. Chem.* **46** 1465–1468.

DEMPSTER, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.* **29** 995–1010.

FINNEY, D. J. (1956). Multivariate analysis and agricultural experiments. *Biometrics* **12** 67–71.

FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugenics* **8** 376–386.

GABBE, J. D. and WILK, M. B. (1960). A study of some factors affecting the quality of television pictures. Unpublished.

GRAYBILL, F. A. (1961). *An Introduction to Linear Statistical Models.* **1** McGraw-Hill, New York.

HOTELLING, HAROLD (1931). The generalization of Student's ratio. *Ann. Math. Statist.* **2** 360–378.

HOTELLING, HAROLD (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.* **24** 417–441 and 498–520.

HOTELLING, HAROLD (1947). Multivariate quality control, illustrated by the air testing of sample bombsights. *Techniques of Statistical Analysis.* 111–184. McGraw-Hill, New York.

LAUE, R. V. (1961). A multivariate approach to screening test data. Unpublished manuscript.

LAWLEY, D. N. (1938). A generalization of Fisher's z test. *Biometrika* **30** 180–187.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India.* **12** 49–55.

NEYMAN, J. and PEARSON, E. S. (1933). On the testing of statistical hypotheses in relation to probabilities *a priori*. *Proc. Cambridge Philos. Soc.* **29** 492–510.

PATNAIK, P. B. (1949). The non-central χ^2 and F-distributions and their approximations. *Biometrika* **36** 202–232.

PEARCE, S. C. (1953). *Field Experimentation with Fruit Trees and Other Perennial Crops.* Commonwealth Bureau of Horticulture and Plantation Crops, East Malling, England.

ROY, J. (1958). Step-down procedure in multivariate analysis. *Ann. Math. Statist.* **29** 1177–1187.

ROY, S. N. (1939). p-statistics or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā* **4** 381–396.

ROY, S. N. and BOSE, R. C. (1953). Simultaneous confidence interval estimation. *Ann. Math Statist.* **24** 513–536.

ROY, S. N. and GNANADESIKAN, R. (1957). Further contributions to multivariate confidence bounds. *Biometrika* **44** 399–410.

ROY, S. N. and GNANADESIKAN, R. (1958). A note on "Further contributions to multivariate confidence bounds." *Biometrika* **45** 581.

ROY, S. N., GNANADESIKAN, R. and SRIVASTAVA, J. N. (1962). Analysis and design of quantitative multiresponse experiments. Unpublished monograph.

SATTERTHWAITE, F. E. (1941). Synthesis of variance. *Psychometrika* **6** 309–316.

SCHEFFÉ, HENRY. (1959). *The Analysis of Variance*. Wiley, New York.

SHEPARD, R. N. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function, I. *Psychometrika* **27** 125–140.

SHEPARD, R. N. (1962b). The analysis of proximities: multidimensional scaling with an unknown distance function, II. *Psychometrika* **27** 219–246.

SMITH, H., GNANADESIKAN, R. and HUGHES, J. B. (1962). Multivariate analysis of variance (MANOVA). *Biometrics* **18** 22–41.

TERRY, M. E. (1955). On the analysis of planned experiments. *National Convention Transactions 1955*. 553–556. Amer. Soc. Qual. Control.

TUKEY, JOHN W. (1962). The future of data analysis. *Ann. Math Statist.* **33** 1–67.

TUKEY, JOHN W. (1953). The problem of multiple comparisons. Unpublished manuscript.

TUKEY, JOHN W. (1949a). One degree of freedom for nonadditivity. *Biometrics* **5** 232–242.

TUKEY, JOHN W. (1949b). Dyadic anova, an analysis of variance for vectors. *Human Biology* **21** 65–110.

WILK, M. B. and GNANADESIKAN, R. (1961). Graphical analysis of multiresponse experimental data using ordered distances. *Proc. Nat. Acad. Sci. U. S. A.* **47** 1209–1212.

WILK, M. B. and GNANADESIKAN, R. (1963). Graphical analysis of variance. Unpublished.

WILK, M. B., GNANADESIKAN, R., and HUYETT, MISS M. J. (1962a). Probability plots for the gamma distribution. *Technometrics* **4** 1–20.

WILK, M. B., GNANADESIKAN, R., and HUYETT, MARILYN J. (1962b). Estimation of parameters of the gamma distribution using order statistics. *Biometrika* **49** 525–545.

WILK, M. B., GNANADESIKAN, R., HUYETT, MISS M. J., and LAUH, MISS E. (1962c). A study of alternate compounding matrices used in a graphical internal comparisons procedure. Unpublished.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24** 471–494.

YATES, F. and HEALY, M. J. R. (1951). Statistical methods in anthropology. *Nature* **168** 1116–1117.