The Annals of Mathematical Statistics
1970, Vol. 41, No. 1, 292–304

USE OF MAXIMUM LIKELIHOOD FOR ESTIMATING ERROR VARIANCE FROM A COLLECTION OF ANALYSIS OF VARIANCE MEAN SQUARES

BY R. GNANADESIKAN AND M. B. WILK

Bell Telephone Laboratories, Incorporated

0. Summary. Given a collection of analysis of variance mean squares, not all of which necessarily have the same degrees of freedom, the present paper describes a method of "mapping" them so as to facilitate the statistical structuring of the mean squares. Even under a null model of no real effects, the mean squares do not have the same distribution because their degrees of freedom may differ, and the ordered mean squares cannot be regarded as the usual order statistics of a sample from a single common distribution.

If the ordered mean squares in a general orthogonal analysis of variance are $0 < S_1 \leqq S_2 \leqq \cdots \leqq S_K$ with corresponding degrees of freedom, $\nu_1, \nu_2, \cdots, \nu_K$, then the inferential reference set in the present approach is one obtained by so-called *complete conditioning*, i.e., repeated sampling from a set of K populations such that the ith ordered mean square will be considered to have come from the population associated with ν_i degrees of freedom, for $i = 1, 2, \cdots, K$.

The approach consists of obtaining from each of the ordered mean squares, in turn, a maximum likelihood estimate of a presumed common error variance based on an order statistics formulation which employs complete conditioning of the mean squares. Methods of obtaining the sequence of maximum likelihood estimates as well as two graphical modes of displaying them are described. Illustrative examples are included.

1. Introduction. The present paper describes procedures to aid in developing an appropriate estimate of error variance from a collection of analysis of variance mean squares. The textbook approach to analysis of variance is to designate an error term on some prior basis (often just assumption) and proceed with formal analyses and comparisons. The error variance is usually obtained from replication, or from block-treatment interactions, or from selected interactions (usually the higher order ones) of treatments. Clearly, the error term plays a very important role in the analysis of variance and its determination should not be left entirely to pre-experimental judgment.

Many of the probability plotting procedures used for internal comparisons in the context of the analysis of variance (cf. Daniel [2], Wilk and Gnanadesikan [4], [5], [6]) do not require a pre-specification of an error term, and the probability plots provide data-determined error configurations against which departures may be sensitively delineated.

Another internal comparisons approach to the problem of bootstrapping an error variance from a collection of mean squares would be to use the background

Received February 26, 1969.

null statistical model that each mean square in the collection reflects just error variance and then to employ an order statistics formulation for estimating the presumed common error variance from each of the ordered mean squares in turn. The sequence of estimates of the error variance thus generated can then be internally studied, preferably graphically, for grouping those mean squares which yield cohesive estimates as belonging to the error configuration. For the case when all the mean squares in the collection have equal degrees of freedom, order statistics formulations for estimating the error variance have been used by Wilk, Gnanadesikan and Freeny [7] and by Wilk, Gnanadesikan and Huyett [8].

The difficulty in the general analysis of variance circumstance involving mean squares with different degrees of freedom is that, even under null assumptions of scaled central χ^2-distributions, the ordered mean squares cannot be regarded as the usual order statistics of a sample from a single common distribution. In the case of unequal components, several alternate conceptions are possible of how to formulate the joint distribution of the ordered observations.

One formulation, called *group conditioning*, has been considered by Wilk et al. [9], for the order statistics from unequal components, having gamma distributions with possibly different known shape parameters but with the same unknown scale parameter. In the present work, the concept used in associating the ordered observations with the parent population is one of *complete conditioning*, which is described in Section 2 and whose cogency for the analysis-of-variance application is discussed elsewhere. (cf. Section 6.)

For reasons of convenience and generality, the development in the subsequent sections is in terms of gamma distributions instead of χ^2 distributions which are special cases of gamma distributions. Section 2 introduces some notation and states the problem, while the method used to determine the maximum likelihood estimate of the presumed common scale parameter from the observed value of an ordered shape-scaled gamma variate is described in Section 3. The use of the estimation procedure for obtaining a sequence of estimates of error variance from ordered analysis of variance mean squares and the utilization of the sequence for statistically structuring the collection of mean squares are discussed in Section 4. Section 5 contains two illustrative numerical examples and Section 6 consists of concluding remarks and discussion. An appendix provides recurrence relations for evaluating the likelihood function.

2. Problem and notation. Let $X_1, X_2, \cdots, X_K$ be a sample of mutually independent random variables, with X_i having a gamma distribution with known shape parameter, η_i^*, and unknown scale parameter, λ, $i = 1, 2, \cdots, K$. The shape-scaled random variable, $S_i^* = X_i/\eta_i^*$, would have the probability density,

$$f(s_i^*; \lambda, \eta_i^*) = [(\lambda\eta_i^*)^{\eta_i^*}/\Gamma(\eta_i^*)] \exp(-\lambda\eta_i^* s_i^*) s_i^{*\eta_i^* - 1}, \quad s_i^* > 0, \lambda > 0, \eta_i^* > 0. \tag{1}$$

If $S_1 \leqq S_2 \leqq \cdots \leqq S_K$ denote the *ordered* shape-scaled quantities and $\eta_1, \eta_2, \cdots, \eta_K$

are the shape parameters associated with the ordered S_i's, then the general joint density of $S_1, \cdots, S_K$ is

$$C \prod_{i=1}^{K} f(s_i; \lambda, \eta_i), \qquad 0 < s_1 \leqq s_2 \leqq \cdots \leqq s_K < \infty, \tag{2}$$

where

$$C^{-1} = \int_0^\infty ds_1 \int_{s_1}^\infty ds_2 \cdots \int_{s_{K-1}}^\infty ds_K \prod_{i=1}^{K} f(s_i; \lambda, \eta_i). \tag{3}$$

The general statistical problem considered here is the estimation of λ from the observed value of S_i, for $i = 1, 2, \cdots, K$, in turn, using the concept of *complete conditioning* for the collection of ordered shape-scaled quantities. That is, at each stage although one employs the observed value of a single shape-scaled quantity, yet the sample space is constrained so that the smallest observation corresponds to the population whose shape parameter is η_1, the next smallest observation corresponds to one whose shape parameter is η_2, etc., and the largest observation corresponds to the population with shape parameter η_K. Geometrically, suppose one visualizes a K-dimensional sample space with coordinates, $C_{\eta_1}, C_{\eta_2}, \cdots, C_{\eta_K}$, having shape-scaled gamma distributions with respective shape parameters $\eta_1, \eta_2, \cdots, \eta_K$ (which correspond to the ordered shape-scaled $S_1, \cdots, S_K$ as observed). By completely conditioned sampling is meant constraining the sample to the conical region, $0 \leqq C_{\eta_1} \leqq C_{\eta_2} \leqq \cdots \leqq C_{\eta_K} < \infty$.

Under such a complete conditioning, the marginal density of S_i would be

$$\begin{aligned} Cf(s_i; \lambda, \eta_i) \int_0^{s_i} \int_0^{s_{i-1}} \cdots \int_0^{s_2} \prod_{j=1}^{i-1} f(s_j; \lambda, \eta_j)\, ds_j \\ \cdot \int_{s_i}^\infty \int_{s_{i+1}}^\infty \cdots \int_{s_{K-1}}^\infty \prod_{j=i+1}^{K} f(s_j; \lambda, \eta_j)\, ds_j. \end{aligned} \tag{4}$$

Equation (4) is obtained by the familiar multinomial argument for the marginal density of the ith order statistic, involving the probabilities of observing $(i-1)$ observations less than the observed ith value, one "near" it, and $(K-i)$ larger than it. The complexity in the formulae for these probabilities, especially for the $(i-1)$ being less and the $(K-i)$ being larger than s_i, is due, in the present case, to both the complete conditioning and the inherent problem of unequal statistical components in the order statistics formulation.

The present approach uses the marginal density of S_i, for $i = 1, 2, \cdots, K$, as the likelihood function of λ, given the observed value s_i of S_i, and the known values of $\eta_1, \cdots, \eta_K$, and obtains a sequence of estimates, $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_K$, of λ where $\hat{\lambda}_i$ corresponds to s_i.

3. Evaluation of the likelihood function. Since the constant, C, does not involve λ, the likelihood function provided by (4) may be rewritten as

$$\begin{aligned} \mathscr{L}_i(\lambda \mid S_i = s_i; \eta_1, \cdots, \eta_K) \propto f_i(s_i) J_{s_i}(T_{i-1}, P_{i-1}; \cdots; T_1, P_1) \\ \times I_{s_i}(T_{i+1}, P_{i+1}; \cdots; T_K, P_K), \end{aligned} \tag{5}$$

where

$$f_i(u) = (T_i)^{P_i+1}[\Gamma(P_i+1)]^{-1}u^{P_i}\exp(-T_i u), \qquad u > 0,\ T_i > 0,\ P_i > -1;$$

$$J_x(T_{i-1}, P_{i-1}; \cdots; T_1, P_1)$$
$$= \int_0^x du_{i-1} \int_0^{u_{i-1}} du_{i-2} \cdots \int_0^{u_2} du_1 \prod_{j=1}^{i-1} f_j(u_j) = J_x(i-1), \quad \text{say};$$

$$I_x(T_{i+1}, P_{i+1}; \cdots; T_K, P_K)$$
$$= \int_x^\infty du_{i+1} \int_{u_{i+1}}^\infty du_{i+2} \cdots \int_{u_{K-1}}^\infty du_K \prod_{j=i+1}^{K} f_j(u_j) = I_x(K-i), \quad \text{say};$$

$T_i = \lambda\eta_i$, $P_i = \eta_i - 1$, and $i = 1, 2, \cdots, K$. Thus, for $i = 1, 2, \cdots, K$ in turn, $\hat{\lambda}_i$ is the value of λ which maximizes $\mathscr{L}_i$ or, equivalently, the value that maximizes

$$L_i(\lambda) = \ln \mathscr{L}_i = \ln f_i(s_i) + \ln J_{s_i}(i-1) + \ln I_{s_i}(K-i). \tag{6}$$

For each i, the likelihood equation, $dL_i/d\lambda = 0$, may be shown to reduce to

$$\begin{aligned} \lambda^{-1} = \eta^{-1}\{\eta_i s_i + \textstyle\sum_{j=1}^{i-1} \eta_j E(U_j \mid U_1 \leqq U_2 \leqq \cdots U_{i-1} \leqq s_i) \\ + \textstyle\sum_{l=i+1}^{K} \eta_l E(U_l \mid s_i \leqq U_{i+1} \leqq \cdots \leqq U_k)\}, \end{aligned} \tag{7}$$

where $\eta = \sum_{i=1}^{K} \eta_i$, U_r is a shape-scaled gamma random variable with density as in Equation (1) having parameter values λ and η_r and where $E(U_j \mid U_1 \leqq U_2 \leqq \cdots \leqq U_{i-1} \leqq s_i)$ and $E(U_l \mid s_i \leqq U_{i+1} \leqq \cdots \leqq U_K)$ denote, respectively, conditional expectations of U_j and U_l under the complete conditioning. These expected values are themselves functions of λ and hence iterative methods would be necessary for solving Equation (7).

However, since L_i involves a single parameter and since $S_K{}^{-1}$ and $S_1{}^{-1}$ directly provide lower and upper bounds for $\hat{\lambda}_i$ for all i, the present authors, rather than solving the likelihood equation iteratively, adopted an approach of directly computing and plotting L_i with a consequent determination of the maximum likelihood estimate itself. For this purpose, at the ith stage ($i = 1, 2, \cdots, K$), methods are needed for computing the $(i-1)$-dimensional integral, $J_{s_i}(i-1)$, and the $(K-i)$-dimensional integral, $I_{s_i}(K-i)$. The recurrence relations given in the appendix provide an inductive scheme, valid when the shape parameters are all integer valued, whereby the integral J of any specified dimension can be computed from a linear combination of J's of one less dimension and with suitably adjusted arguments. Results are given in the appendix which provide a similar inductive scheme for evaluating the integrals, I, again when all the η_i's are integers. The appendix also proposes a method for approximate evaluation of I and J when some of the shape parameters involved are such that twice their values are integers while the others are integer valued. In the analysis of variance framework, wherein the S_i's would correspond to ordered mean squares, the situation when the η_i's are all integers corresponds to the special case when the degree of freedom of every mean square is even. The situation when some η_i's are integers while others are such that twice their values are integers corresponds to the general case when the degrees of freedom of the different mean squares are either even or odd.

Thus, given the observed values of S_i and $\eta_1, \cdots, \eta_K$, one can compute and graph the log likelihood function, $L_i(\lambda)$, and determine $\hat{\lambda}_i$ as that value of λ for which $L_i(\lambda)$ is a maximum. For explicitly determining the value of $\hat{\lambda}_i$, one can use the fact that it lies in the interval, $[S_K{}^{-1}, S_1{}^{-1}]$, and successively halve the intervals in which $\hat{\lambda}_i$ is verified to lie, until $\hat{\lambda}_i$ is determined with the desired accuracy.

A measure of the sharpness of the ith likelihood function in the neighborhood of its maximum, which may also be thought of as an estimate of the asymptotic variance[1] of the maximum likelihood estimate, is provided by $-1/(d^2L_i/d\lambda^2)_{\hat{\lambda}_i}$. Using second-order differences to approximate second-order derivatives, the sharpness of the likelihood is measured by

$$A_i^2 = h_i^2/[2\hat{L}_i - L_i^{(1)} - L_i^{(2)}], \tag{8}$$

where

h_i = distance from $\hat{\lambda}_i$ to $\lambda^{(1)}$ where $\lambda^{(1)}$ is that observed value $S_j{}^{-1}$, $j = 1, \cdots, K$, which is closest to $\hat{\lambda}_i$,

$\hat{L}_i = L_i(\hat{\lambda}_i)$ = maximum value of $L_i(\lambda)$,

$L_i^{(1)} = L_i(\lambda^{(1)})$,

$L_i^{(2)} = L_i(\lambda^{(2)})$, $\lambda^{(2)} = 2\hat{\lambda}_i - \lambda^{(1)}$.

4. Sequenced estimation of scale parameter as a basis for statistically structuring a collection of mean squares. The maximum likelihood estimation procedure described in the last section may be used with a collection of analysis of variance mean squares to generate a sequence of estimates of a presumed common error variance, σ^2. Thus, suppose $0 < S_1 < S_2 \leqq \cdots \leqq S_K \leqq \infty$ denote the ordered mean squares with corresponding degrees of freedom, $\nu_1, \nu_2, \cdots, \nu_K$. Under the null statistical model, which is used merely to develop the methodology that can be employed for studying the departures from the assumptions, the sums of squares corresponding to the unordered mean squares are considered to be distributed as $\sigma^2\chi^2$ variables with specified degrees of freedom. In this null view, what ties the mean squares together statistically is that each of them is considered to be an estimate of σ^2. What is sought here is a transformation of the ordered mean squares which will make allowance for the order relationships amongst, and for the statistical conditioning of, the mean squares and will provide a basis for appreciation of departures from the null model.

In the viewpoint of the present paper, the observed order relationship amongst the mean squares is considered to be relevant information in determining the appropriate reference set for statistical inference. This implies inferential considerations relative to repeated sampling from a set of K populations such that the ith ordered mean square will be constrained to have come from the population associated with

[1] The theory of maximum likelihood estimation based on unequal statistical components, such as in the present application, is yet to be developed and this statement is a suggestive and interpretive one and not a mathematically formal one.

ν_i degrees of freedom, for $i = 1, 2, \cdots, K$. In other words, complete conditioning is imposed.

Correspondence of this problem involving analysis of variance mean squares with the one considered in Sections 2 and 3 can be achieved by considering the X_i's of Section 2 to be the sums of squares divided by 2, λ to be σ^{-2} and $\eta_i = \nu_i/2$, for $i = 1, 2, \cdots, K$. The log likelihood function, which is now considered as a function of σ^2, may be obtained from Equation (6) with s_i now denoting the observed value of the ith ordered mean square S_i, and with the arguments of the integrals J and I being defined by $T_i = \nu_i/2\sigma^2$ and $P_i = (\nu_i/2) - 1$, for $i = 1, 2, \cdots, K$. Equation (7) reduces to

$$\begin{aligned}\sigma^2 = \nu^{-1}\{\nu_i s_i + \textstyle\sum_{j=1}^{i-1} \nu_j E(U_j \mid U_1 \leqq \cdots \leqq U_{i-1} \leqq s_i) \\ + \textstyle\sum_{l=i+1}^{K} \nu_l E(U_l \mid s_i \leqq U_{i+1} \leqq \cdots \leqq U_K)\},\end{aligned} \tag{9}$$

where $\nu = \sum_{i=1}^{K} \nu_i$. This form of the likelihood equation is intuitively reasonable in that it provides a "pooled estimate", in which the mean square in focus (namely the ith one) is used explicitly while the other mean squares are replaced in the pooling by their expected values under complete conditioning. As before, however, the conditional expected values in Equation (9) are themselves functions of σ^2, and the approach to be adopted is not iterative solution but evaluation and graphing of $L_i(\sigma^2)$ and thence a determination of $\hat{\sigma}_i^2$, for $i = 1, 2, \cdots, K$ in turn.

The approximate measure, provided by Equation (8), of the estimated asymptotic variance of $\hat{\sigma}_i^2$ is,

$$A_i^2 = h_i^2/[2\hat{L}_i - L_i^{(1)} - L_i^{(2)}], \tag{10}$$

where

$h_i =$ distance from $\hat{\sigma}_i^2$ to the nearest observed ordered mean square, say S,

$$\hat{L}_i = L_i(\hat{\sigma}_i^2),$$

$$L_i^{(1)} = L_i(S),$$

$$L_i^{(2)} = L_i(2\hat{\sigma}_i^2 - S).$$

Corresponding to the ordered mean squares $S_1 \leqq S_2 \leqq \cdots \leqq S_K$, one can thus obtain a sequence of estimates, $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \cdots, \hat{\sigma}_K^2$, with respective statistical allowances, $A_1, A_2, \cdots, A_K$. Although the mean squares are ordered the estimates need not and, in general, will not be. If, however, the smaller mean squares reflect only error variance while the larger ones reflect additional real variation, then one expects that the σ^2 estimates obtained from the smaller mean squares would be relatively cohesive while those derived from the larger mean squares would be distinctly bigger. Intuitively, one might expect that the "break" in the sequence of estimates, say at S_m, would persist in the estimates obtained from each S_j for $j > m$.

One graphical presentation of the results of the sequenced estimation is to plot, in a single figure, all the ratios of the log likelihood functions, $L_1(\sigma^2)/L_1(\hat{\sigma}_1^2), \cdots,$

$L_K(\sigma^2)/L_K(\hat{\sigma}_K^2)$, each in the neighborhood of the corresponding maximum likelihood estimate of σ^2. In such a picture one could see both jumps in the sequence of estimates and the differing sharpness of the various likelihood functions. The shifts in the location of the likelihood function would be caused by the mean square in question reflecting a possibly real effect, and the sharpness of the likelihood function depends both on the ordering of the mean square and on its degrees of freedom.

Another graphical display would be provided by plotting the estimate of σ^2 against the mean square in focus, i.e., a plot of the K points, $(S_i, \hat{\sigma}_i^2)$ for $i = 1, 2, \cdots, K$. In such a plot one could also indicate statistical allowances for each estimate by a band of $\pm 2A_i$ for $\hat{\sigma}_i^2$ $(i = 1, \cdots, K)$. Examples of both types plots are given in the next section.

5. Examples. Two sets of data are used to illustrate the methods described heretofore. The first example is based on a set of eight computer-generated mean squares: two with 2 degrees of freedom, two with 4, two with 6 and two with 8 degrees of freedom. The value of σ^2 used in the process was unity. Table I shows the ordered mean squares with the corresponding degrees of freedom, the maximum likelihood estimates obtained from each ordered mean square in turn using the completely conditioned formulation described earlier, and the values of the statistical allowances, A_i's.

TABLE I
Augmented analysis of variance table
(Monte Carlo Example; $K = 8$, $\sigma^2 = 1$)

i	Ordered mean squares (S_i)	d.f.	$\hat{\sigma}_i^2$	A_i
1	0.1233	2	0.5847	0.4353
2	0.4514	6	1.1421	0.5096
3	0.5386	2	1.0036	0.3816
4	0.5759	4	0.8298	0.2572
5	1.0378	4	1.2024	0.3487
6	1.1018	8	1.0459	0.2761
7	1.3208	6	1.0186	0.2643
8	1.6340	8	0.9784	0.2669

In this null example, except for $\hat{\sigma}_1^2$ being noticeably small, the different estimates of σ^2 are seen to be reasonably concordant and are all within $\pm 2A$ of the true value of $\sigma^2 = 1$, as well as of each other. Figure 1 shows on a single graph plots of the ratios, $L_i(\sigma^2)/L_i(\hat{\sigma}_i^2)$, for a representative choice of values of $i = 1, 3, 5$ and 7. A reasonable concordance of the estimates is evident. Also, as one might expect, the plots corresponding to the mean squares with larger degrees of freedom tend to be sharper than those for smaller degrees of freedom.

Another presentation of the information is given in Figure 2, which shows a plot

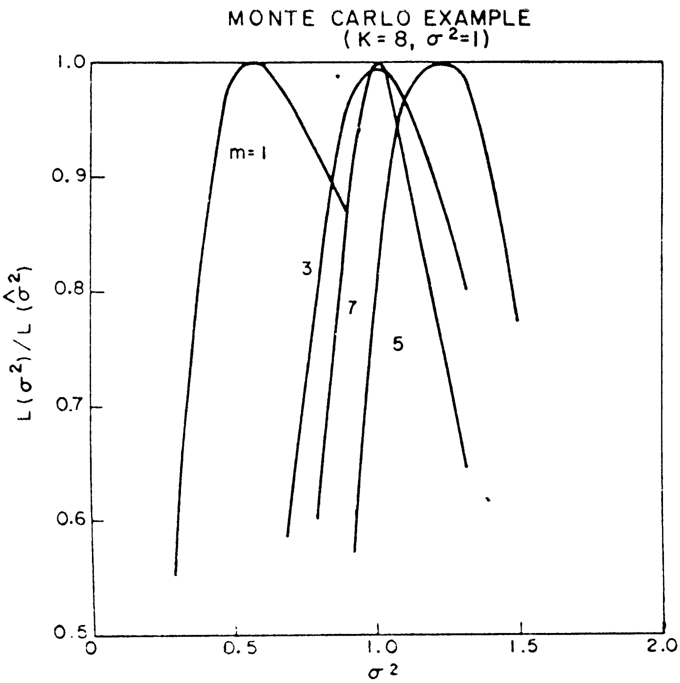
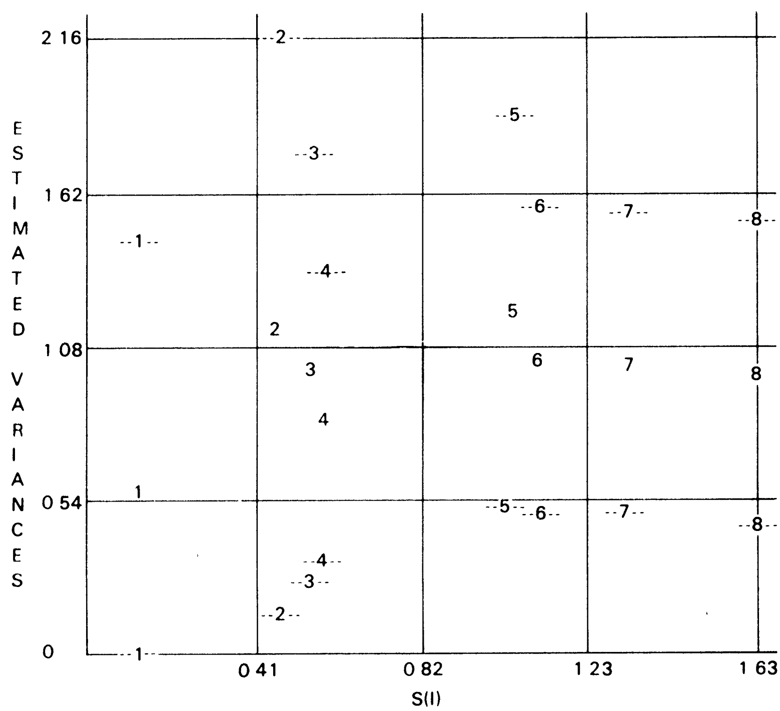


FIG. 1.



PLOT OF ESTIMATED VARIANCES VS ORDERED MEAN SQUARES
MONTE CARLO EXAMPLE (K = 8, (SIGMA)^2 = 1)
--J-- REPRESENTS THE CORRESPONDING UPPER AND LOWER
BOUNDS (ESTIMATE $\pm$2A POINTS)

FIG. 2.

of the estimates, $\hat{\sigma}_i^2$, against the ordered mean squares, S_i, with bands about the estimates corresponding to $\pm$ twice the allowances, A_i.[2] In this configuration one can see that $\hat{\sigma}_1^2$ is low but not unduly so in terms of its allowance for variation and also that, in spite of noticeable breaks between certain of the observed ordered null mean squares (e.g., between S_4 and S_5), the estimates themselves are quite concordant.

The second example, taken from Bennett and Franklin ([1] pages 542–545), deals with six analysis of variance mean squares from the analysis of a 3×5 factorial experiment for studying the effects of 3 types of oil on the wear of 5 piston rings and involving 5 replicate observations.

Table II shows the ordered mean squares identified by source and by the degrees of freedom. Also given in Table II are the maximum likelihood estimates, $\hat{\sigma}_i^2$, and the corresponding statistical allowances, A_i, for $i = 1, 2, \cdots, 6$.

TABLE II
Augmented analysis of variance table for a 3×5 factorial experiment
(Bennett & Franklin, 1954, pp. 542–545)

i	Source	Ordered mean squares (S_i)	d.f.	$\hat{\sigma}_i^2$	A_i
1	Oils × rings	.005606	8	.00836	.0022
2	Residual	.006061	48	.00684	.0012
3	Tests (wn. oils)	.024932	8	.0239	.0048
4	Replications	.035151	4	.0278	.0065
5	Oils	.069123	2	.0431	.0120
6	Rings	1.213747	4	.559	.0180

The last estimate, $\hat{\sigma}_6^2$, corresponding to the largest mean square, is clearly much larger than the rest of the estimates and, in fact, it corresponds to a real effect as claimed by Bennett and Franklin [1] who use an F-test for establishing statistical significance at the 0.1 % level.

Figure 3 shows plots of $L_i(\sigma^2)/L_i(\hat{\sigma}_i^2)$, for $i = 1, 2, \cdots, 6$, for this example. In order to show $L_6(\sigma^2)/L_6(\hat{\sigma}_6^2)$ on the same graph as, but without loss of information concerning, the others, the abscissa scale for it is indicated separately by the horizontal scale at the top of the figure.

Figure 4 provides a graphical summary by plotting the estimates, with bands based on the corresponding allowances, against the ordered mean squares. The "largest" point, $(S_6, \hat{\sigma}_6^2)$, which is clearly separable from the remaining ones, is omitted in Figure 4 so that the configuration of the "smaller" points may be studied more easily.

Both plots suggest a further possible partitioning of the mean squares into two

[2] Since $\sigma_1^2 - 2A_1 < 0$, the lower limit for the first estimate is shown as zero in Figure 2.

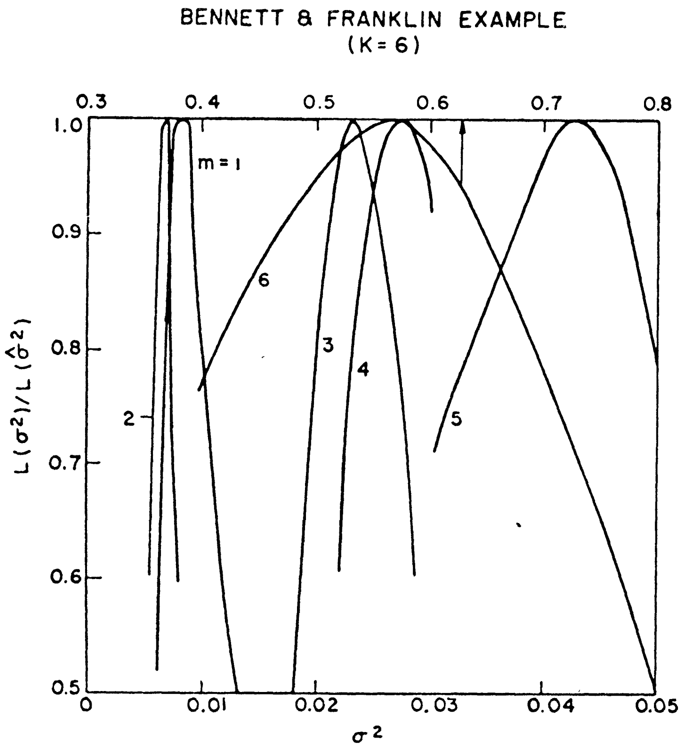
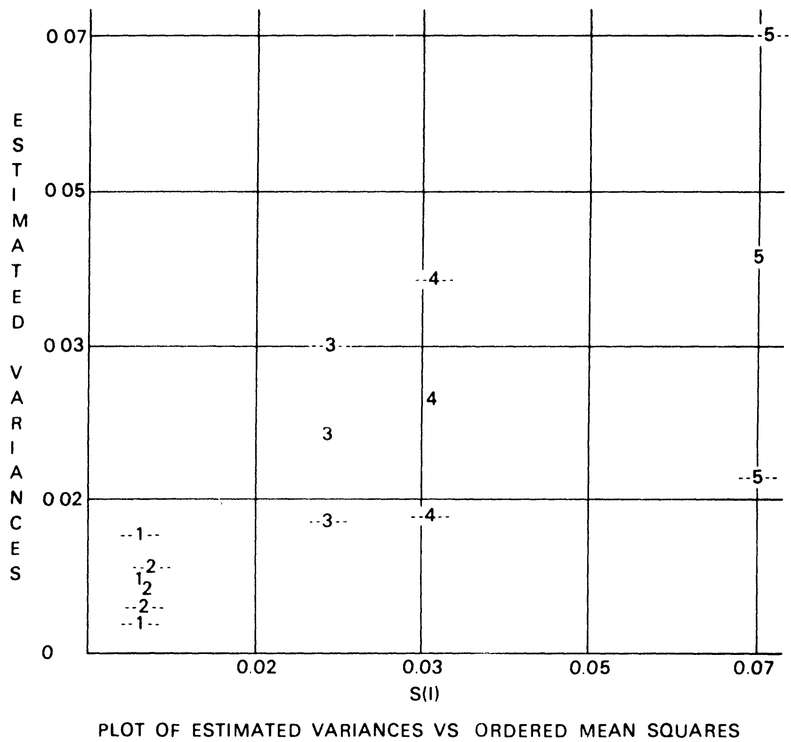


Fig. 3.



PLOT OF ESTIMATED VARIANCES VS ORDERED MEAN SQUARES
MEAN SQUARES FROM 3x5 FACTORIAL EXPERIMENT
--J-- REPRESENTS THE CORRESPONDING UPPER AND LOWER
BOUNDS (ESTIMATE ±2A POINTS)

Fig. 4.

groups, one consisting of S_1 and S_2 and the other consisting of S_3, S_4 and S_5, although the internal cohesiveness of the corresponding estimates of σ^2 is not very stably inferrable because of the small numbers of mean squares involved.

6. Discussion. The procedure described in this paper is intended as a method for augmenting the usual analysis of variance table. Primarily the concern is with exhibiting the information to facilitate comparisons rather than with providing formal probabilistic inferences. As an informal internal comparisons procedure, the method is based on the unifying statistical notion of a common error variance underlying each of the related mean squares. This statistical strawman, the null hypothesis or viewpoint, is deliberately set up to provide a basis for developing techniques to lead to its critical evaluation.

Clearly, the particular estimates, the $\hat{\sigma}_i^2$, depend upon the number, K, included in the supposed "null" collection of mean squares. When one or more of these are "tagged" as definitely not being solely associated with error, the remaining mean squares need to be reanalyzed. This is similar to the need to replot after omission of "discrepant" points in probability plotting. Also, based on the indications of a first-cut analysis of a collection of mean squares employing the method described in Section 4, one may be able to infer a relatively homogeneous group of "error" mean squares, and subsequent analyses might then be directed towards improving the statistical properties of the estimate of error variance by appropriate pooling, etc.

The argument for using the concept of complete conditioning as the appropriate one for the present application is the same as that for the method of generalized probability plotting described by Gnanadesikan and Wilk [3]. A drawback of the present method is the heavy computation involved in the recursion scheme whose use is predicated on the availability of modern high-speed computers. The group conditioning approach (see Wilk, et al. [9]) involves substantially less computing load and seems often to give somewhat comparable results.

Acknowledgments. The authors are grateful to Miss Elizabeth Lauh for programming the computational scheme for evaluating the log likelihood function and for plotting the ratios, $L_i(\sigma^2)/L_i(\hat{\sigma}_i^2)$. They also wish to thank T. L. DeChaine for his help in obtaining the computer plots of the points, $(S_i, \hat{\sigma}_i^2)$, with the bands of $\pm 2A_i$.

APPENDIX ON EVALUATION OF THE INTEGRALS I AND J

Recall the definitions,

$$J_x(i-1) = J_x(T_{i-1}, P_{i-1}; \cdots; T_1, P_1)$$
$$= \int_0^x du_{i-1} \int_0^{u_{i-1}} du_{i-2} \cdots \int_0^{u_2} du_1 \prod_{j=1}^{i-1} f_j(u_j),$$

and

$$I_x(K-i) = I_x(T_{i+1}, P_{i+1}; \cdots; T_K, P_K)$$
$$= \int_x^\infty du_{i+1} \int_{u_{i+1}}^\infty du_{i+2} \cdots \int_{u_{K-1}}^\infty du_K \prod_{j=i+1}^{K} f_j(u_j),$$

where

$$f_j(u) = (T_j)^{P_j+1}[\Gamma(P_j+1)]^{-1}u^{P_j}\exp(-T_j u), \qquad u > 0,\ T_j > 0,\ P_j > -1.$$

Assume, in addition, that P_j, for all j, is a nonnegative integer. Then, repeated integrations by parts, yield the following results for special cases:

(1a)
$$\begin{aligned} J_x(1) = J_x(T_1, P_1) &= \textstyle\int_0^x f_1(u_1)\,du_1 \\ &= T_1^{P_1+1}[\Gamma(P_1+1)]^{-1}\textstyle\int_0^x u_1^{P_1}\exp(-T_1 u_1)\,du, \\ &= 1-\exp(-T_1 x)\textstyle\sum_{j=0}^{P_1}\alpha_j, \end{aligned}$$

where $\alpha_0 = 1$ and $\alpha_j = j^{-1}(T_1 x)\alpha_{j-1}$, for $j = 1, \cdots, P_1$.

(1b)
$$\begin{aligned} I_x(1) = I_K(T_K, P_K) &= \textstyle\int_x^\infty f_K(u_K)\,du_K \\ &= (T_K)^{P_K+1}[\Gamma(P_K+1)]^{-1}\textstyle\int_x^\infty u_K^{P_K}\exp(-T_K u_K)\,du_K \\ &= \exp(-T_K x)\textstyle\sum_{j=0}^{P_K}\beta_j, \end{aligned}$$

where $\beta_0 = 1$ and $\beta_j = j^{-1}(T_K x)\beta_{j-1}$, for $j = 1, 2, \cdots, P_K$.

(2a)
$$\begin{aligned} J_x(2) &= J_x(T_2, P_2;\ T_1 P_1) \\ &= \frac{T_1^{P_1+1}}{\Gamma(P_1+1)}\frac{T_2^{P_2+1}}{\Gamma(P_2+1)}\textstyle\int_0^x u_2^{P_2}\exp(-T_2 u_2)\,du_2\int_0^{u_2} u_1^{P_1}\exp(-T_1 u_1)\,du_1 \\ &= J_x(T_2, P_2) - \left(\frac{T_2}{T_1+T_2}\right)^{P_2+1}\textstyle\sum_{j=0}^{P_1}\alpha_j J_x(T_2+T_1, P_2+j), \end{aligned}$$

where $\alpha_0 = 1$, $\alpha_j = [T_1/(T_1+T_2)][(P_2+j)/j]\alpha_{j-1}$ for $j = 1, \cdots, P_1$.

Each one-dimensional J-integral on the right-hand side may itself be evaluated by employing the formula (1a) above.

(2b)
$$\begin{aligned} I_x(2) &= I_x(T_{K-1}, P_{K-1};\ T_K, P_K) \\ &= \left\{\frac{T_{K-1}^{P_{K-1}+1}}{\Gamma(P_{K-1}+1)}\frac{T_K^{P_K+1}}{\Gamma(P_K+1)}\right\}\textstyle\int_x^\infty u_{K-1}^{P_{K-1}}\exp(-T_{K-1}u_{K-1})\,du_{K-1} \\ &\quad\cdot\textstyle\int_{u_{K-1}}^\infty u_K^{P_K}\exp(-T_K u_K)\,du_K \\ &= \left(\frac{T_{K-1}}{T_{K-1}+T_K}\right)^{P_{K-1}+1}\textstyle\sum_{j=0}^{P_K}\beta_j I_x(T_{K-1}+T_K, P_{K-1}+j), \end{aligned}$$

where $\beta_0 = 1$ and $\beta_j = [T_K/(T_K+T_{K-1})][(P_{K-1}+j)/j]\beta_{j-1}$ for $j = 1, \cdots, P_K$.

Each one-dimensional I-integral in the sum on the right-hand side would itself be computed by employing the scheme in (1b) above.

More generally, when the P_j's are all nonnegative integers, we have the following recursion-reduction formulae:

(3a)
$$\begin{aligned} J_x(i-1) &= J_x(T_{i-1}, P_{i-1};\cdots;\ T_1, P_1) \\ &= J_x(T_{i-1}, P_{i-1};\cdots;\ T_2 P_2) - \left(\frac{T_2}{T_1+T_2}\right)^{P_2+1} \\ &\quad\cdot\textstyle\sum_{j=0}^{P_1}\alpha_j J_x(T_{i-1}, P_{i-1};\cdots;\ T_3 P_3;\ T_2+T_1, P_2+j), \end{aligned}$$

where $\alpha_0 = 1$ and $\alpha_j = [T_1/(T_1+T_2)][(P_2+j)/j]\alpha_{j-1}$ for $j = 1, 2, \cdots, P_1$.

(3b) $$I_x(K-i) = I_x(T_{i+1}, P_{i+1}; \cdots; T_K, P_K)$$
$$= \left(\frac{T_{K-1}}{T_{K-1}+T_K}\right)^{P_{K-1}+1}$$
$$\cdot \sum_{j=0}^{P_k} \beta_j I_x(T_{i+1}, P_{i+1}; \cdots; T_{K-2}, P_{K-2}; T_{K-1}+T_K, P_{K-1}+j),$$

where $\beta_0 = 1$ and $\beta_j = [T_K/(T_K+T_{K-1})][(P_{K-1}+j)/j]\beta_{j-1}$ for $j = 1, 2, \cdots, P_K$.

Suppose some of the P_j's, namely $P_{j_1}, P_{j_2}, \cdots, P_{j_r}$, are not integers but, for each i, $2P_{j_i}$ is an integer. Then an approximate interpolation procedure is to use the average of two calculations. In one of these, use the arguments $(P_{j_i}+\frac{1}{2})$ in place of P_{j_i}, $i = 1, 2, \cdots, r$. In the other, use $(P_{j_i}-\frac{1}{2})$. This procedure is computationally more practical than a full r-dimensional linear interpolation procedure which would require 2^r separate evaluations.

In the case of analysis of variance mean squares, the P_j's are necessarily either integers or half-integers.

REFERENCES

[1] Bennett, C. A. and Franklin, N. L. (1954). *Statistical Analysis in Chemistry and the Chemical Industry*. Wiley, New York.

[2] Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1** 311–341.

[3] Gnanadesikan, R. and Wilk, M. B. (1970). A probability plotting procedure for general analysis of variance. To appear in *J. Roy. Statist. Soc. Sec. B* **32**.

[4] Wilk, M. B. and Gnanadesikan, R. (1961). Graphical analysis of multiresponse experimental data using ordered distances. *Proc. Nat. Acad. Sci. USA* **47** 1209–1212.

[5] Wilk, M. B. and Gnanadesikan, R. (1964). Graphical methods for internal comparisons in multiresponse experiments. *Ann. Math. Statist.* **35** 613–631.

[6] Wilk, M. B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55** 1–17.

[7] Wilk, M. B., Gnanadesikan, R. and Freeny, Anne E. (1963). Estimation of error variance from smallest ordered contrasts. *J. Amer. Statist. Assoc.* **58** 152–160.

[8] Wilk, M. B., Gnanadesikan, R. and Huyett, Marilyn J. (1963). Separate maximum likelihood estimation of scale or shape parameters of the gamma distribution using order statistics. *Biometrika* **50** 217–221.

[9] Wilk, M. B., Gnanadesikan, R. and Lauh, Elizabeth (1966). Scale parameter estimation from the order statistics of unequal gamma components. *Ann. Math. Statist.* **37** 152–176.