SIMULTANEOUS TESTS OF LINEAR HYPOTHESES AND CONFIDENCE INTERVAL ESTIMATION

BY M. C. VERMA AND M. N. GHOSH

Institute of Agricultural Research Statistics

SECTION 1

IN applied statistics one has to use the same set of data for various kinds of statistical inferences, to test for various types of relations between the variables which are suggested either through theoretical implications or previous experience with similar type of data. Different such statistical tests are not independent in the probability sense and the levels of significance for each of them individually do not provide for the corresponding amount of protection against spurious inferences. The idea of protection level was developed by Tukey for such situations. Ghosh (1955) has introduced the notion of simultaneous level of significance which denotes the probability of rejecting at least one of the null hypotheses $H_1, H_2, \ldots$ etc. when all of them are in fact true. This idea corresponds to the idea of the simultaneous confidence interval estimation of all linear functions of a number of parameters developed by Scheffé (1953) and Roy and Bose and (1953). We shall consider here simultaneous tests of certain linear hypotheses and the related problems of finding simultaneous confidence intervals for groups of parameters. All these tests are based on the least square estimates of the corresponding parameters which has the important property of being quasi-independent as defined by Ghosh (1955). Two tests T_1 and T_2 of hypotheses H_1 and H_2, where H_1 and H_2 do not involve any common parameters, are called quasi-independent when in the test T_1 of H_1, the first and second kinds of error do not involve the parameters of the hypothesis H_2 and *vice versa*.

We shall find here simultaneous confidence intervals for different groups of parameters in a linear set-up and show that they are shorter than those obtained by Scheffé's method. In many problems such groups of parameters are naturally occurring in the problem and does not put any restriction on the problem. The computations involved in the use of these methods are also of the same order as in the use of Scheffé's method.

SECTION 2. SIMULTANEOUS TESTS OF LINEAR HYPOTHESES

Consider the linear model

$$E(y_i) = a_{i1}p_1 + \ldots + a_{im}p_m \qquad (i = 1, \ldots, N) \qquad (2.1)$$

y_i being independent normal r.v.'s with unknown variance σ^2 and $p_1 \ldots p_m$ unknown parameters. Let rank $(a_{ij}) = N_0$ and let $\pi_1 \ldots \pi_R$ be estimable linear functions of the parameters $p_1 \ldots p_m$

$$\pi_i = l_{i1}p_1 + \ldots + l_{im}p_m \qquad (i = 1, \ldots, R)$$

such that the coefficient vectors $(l_{i1} \ldots l_{im})$ form a vector space of rank $R \leq N_0$. Suppose we are interested in the multiple (linear) hypotheses:

$$H_1 : \pi_1 = 0 \quad \ldots\ldots\ldots\ldots \quad \pi_{k_1} = 0$$

$$\vdots$$

$$H_s : \pi_{k_1+\ldots k_{s-1}+1} = 0, \quad \ldots \quad \pi_{k_1+\ldots k_s} = 0 \quad \sum_1^s k_i = R$$

For the simultaneous test of these 's' hypotheses we want that the first kind of error, *i.e.*, the simultaneous level of significance is $\leqslant \alpha$ and that the test has good properties against certian class of alternatives which depend on experience and interest of the experimenter.

Let

$$Y_1 \ldots Y_{k_1}; \quad Y_{k_1+1} \ldots Y_{k_1+k_2}; \ldots Y_{k_1+\ldots k_s}$$

be the best linear estimates of these parameters obtained by the method of least squares. We shall sometimes denote the coefficient vectors of these linear functions by the same symbol, so that we have the alternative notation for the linear function $Y_i = (Y_i, y)$ where (Y_i, y) is the scalar product of the vector Y_i and the observation vector y. From the Markoff's theorem we have an independent estimate of error variance S_e^2 say, with $n_e d.f.$ which is independent of the parameters $p_1, \ldots, p_m$. Then it is obvious that if the test of the hypothesis H_n is based on the linear functions

$$(Y_{b_{n-1}+1}, y), \ldots, (Y_{b_n}, y) \quad b_n = k_1 + \ldots + k_n$$

whose expectations are $\pi_{b_{n-1}+1} \ldots \pi_{b_n}$ respectively, then it will be quasi-independent of the rest of the hypotheses.

Let $U_{b_{n-1}+1} \ldots U_{b_n}$ be ortho-normal vectors forming a basis of the vector space formed by $Y_{b_{n-1}+1} \ldots Y_{b_n}$. Then (U_i, y) is a linear form in (Y_j, y) $(i, j = b_{n-1}+1, \ldots, b_n)$ and

$$E(U_i, y) = \sum_j a_j E(Y_j, y) = \Sigma a_j \pi_j = \phi_i \text{ say} \tag{2.2}$$

On the hypothesis H_n,

$$\frac{\chi_n^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_j [(U_j, y) - E(U_j, y)]^2$$

has a χ^2 distribution with k_n d.f. The second kind of error for this test of H_n shall depend only on the parameters $\pi_{b_{n-1}+1} \ldots \pi_{b_n}$.

The case when the estimates

$$Y_1 \ldots Y_{k_1};\ Y_{k_1+1} \ldots Y_{k_1+k_2}$$

belonging to different hypotheses are orthogonal, has been considered by Ghosh (1955) and later elaborated by Ramachandran (1956). Ramachandran has further provided tables for the special case of $k_1 = k_2$ (2 hypotheses). Nair (1948) has studied the very special case of $k_1 = k_2 = \ldots = k_s = 1$ and provided tables while Ramachandran has elaborated the use of what he calls studentized χ^2 in this connection.

For convenience we consider the non-orthogonal case involving two hypotheses. Ghosh (1955) has shown that in this case there exists no non-singular transformation by which the linear functions can be transformed into mutually orthogonal sets.

Let $U_1 \ldots U_{k_1};\ U_{k_1+1} \ldots U_{k_1+k_2}$ be an orthogonal basis of the vector space formed by $Y_1 \ldots Y_{k_1+k_2}$ so that $U_1 \ldots U_{k_1}$ is a basis of the vector space formed by the first k_1 vectors $Y_1 \ldots Y_{k_1}$ of H_1. Let $E(U_i, y) = \phi_i$ where ϕ_i's are linear functions of $\pi_1 \ldots \pi_{k_1};\ \pi_{k_1+1} \ldots \pi_{k_1+k_2}$ and let $\tilde{Y}_i$ be the normalized vector corresponding to Y_i, *i.e.*, $\tilde{Y}_i = Y_i/|Y_i|$ where $|Y_i|$ is the norm of the vector Y_i and $\tilde{\pi}_i = \pi_i/|Y_i|$.

For convenience we shall assume that $Y_1 \ldots Y_{k_1}$ are mutually orthogonal as also $Y_{k_1+1} \ldots Y_{k_1+k_2}$ but not between the two sets. This may involve a transformation of the parameter set of the problem which leaves the problem essentially invariant.

The relation between U-vectors and y-vectors is expressed by

$$\begin{matrix} k_1 \\ k_2 \end{matrix} \begin{pmatrix} \bar{Y}_1 \\ Y_{\text{II}} \end{pmatrix} = \begin{pmatrix} a & 0 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} U_{\text{I}} \\ U_{\text{II}} \end{pmatrix} \tag{2.3}$$

where $\alpha\,(k_1 \times k_1)$ is an orthogonal matrix, $\gamma\,(k_2 \times k_2)$ is non-singular and $\beta\,(k_2 \times k_1)$ is non-null, since vectors $\tilde{Y}_{\text{I}}$ and $\tilde{Y}_{\text{II}}$ are not mutually orthogonal, and the row vectors of (β, γ) are of unit length

where

$$\tilde{Y}_{\text{I}} = \begin{pmatrix} \tilde{Y}_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{Y}_{k_1} \end{pmatrix} \text{etc.}$$

It is to be noted that an orthogonal transformation of (2.3) leaves the rank of β and orthogonality of α unaltered.

On the basis of (2.3) Ghosh derived 3 different methods of finding confidence regions for the parameters $\tilde{\pi}_1 \ldots \tilde{\pi}_{k_1};\ \tilde{\pi}_{k_1+1} \ldots \tilde{\pi}_{k_1+k_2}$. Method (*a*) is essentially the same as that of Scheffé and Roy-Bose, method (*b*) gives confidence regions for $\pi_1 \ldots \pi_{k_1}$ and $\pi_{k_1+1} \ldots \pi_{k_1+k_2}$, separately while method (*c*) employs a singular transformation of $\begin{pmatrix} Y_{\text{I}} \\ Y_{\text{II}} \end{pmatrix}$ to arrive at sets of linear functions which are mutually orthogonal. We shall develop a modified form of method (*b*) and a new method (*d*) and compare them with the Scheffé's method (*a*) in respect of lengths of confidence intervals. It will be shown that the method (*d*) invariably gives shorter confidence intervals than method (*a*).

Method (*a*).—Along with the joint test of significance for H_1 and H_2 it also gives the confidence intervals for all linear functions of the parameters $\pi_1 \ldots \pi_{k_1};\ \pi_{k_1+1} \ldots \pi_{k_1+k_2}$ with a joint confidence coefficient.

From (2.3) after application of Schwartz's inequality we get

$$|(\tilde{Y}_n, y) - \tilde{\pi}_n|^2 \leqslant \sum_1^{k_1+k_2} [(U_j, y) - \phi_j]^2 \qquad n = 1, \ldots, k_1 + k_2$$

Hence

$$P_r\,[|(\tilde{Y}_n, y) - \tilde{\pi}_n| \leqslant \delta \text{ for all } n = 1, \ldots, k_1 + k_2]$$

$$\geqslant P_r\,[\sum_1^{k_1+k_2} \{(U_j, y) - \phi_j\}^2 \leqslant \delta^2]. \tag{2.4}$$

If

$$P_r\,[\,\Sigma\,\{(U_j, y) - \phi_j\}^2 \leqslant \delta^2]$$

is fixed at $1 - \alpha$, then with confidence coefficient $1 - \alpha$ we get the set of simultaneous linear intervals;

$$(\tilde{Y}_n, y) - \delta_a \leqslant \tilde{\pi}_n \leqslant (\tilde{Y}_n, y) + \delta_a \text{ for all } n = 1, \ldots, k_1 + k_2. \tag{2.5}$$

For the joint test of H_1 and H_2 we may use

$$\chi^2 = \frac{1}{\sigma^2} \sum \{(U_j, y)\}^2$$

and test against $n_e S_e^2/\sigma^2$; when δ^2_a will be found to be equal to

$$(k_1 + k_2) F_{1-a,\, k_1+k_2,\, n_e} \cdot S_e^2.$$

If we want to specify also as to which hypothesis out of H_1 and H_2 is to be accepted (rejected), we may use two test statistics

$$T_1 = \sum_1^{k_1} \frac{(\hat{Y}_n, y)^2}{n_e S_e^2}$$

and

$$T_2 = \sum_{k_1+1}^{k_1+k_2} \frac{(\tilde{Y}_n, y)^2}{n_e S_e^2}.$$

From (2.3) we shall find that if $\pi_1 = 0, \ldots = \pi_{k_1+k_2} = 0$,

$$T_1 (n_e S_e^2) \leqslant \sum_1^{k_1+k_2} (U_j, y)^2$$

and

$$T_2 (n_e S_e^2) \leqslant \sum_1^{k_1+k_2} (U_j, y)^2. \tag{2.6}$$

Hence statistics T_1 and T_2 will provide quasi-independent tests of hypotheses H_1 and H_2 respectively and the upper bound of simultaneous significance level will be provided by the distribution of

$$\sum_1^{k_1+k_2} (U_j, y)^2 / n_e S_e^2$$

so that the critical limits for T_1 and T_2 are the same, *viz.*,

$$\lambda = \frac{k_1 + k_2}{n_e} \cdot F_{1-a,\, k_1+k_2,\, n_e}.$$

If T_i $(i = 1, 2) > \lambda$, we may conclude that H_i is rejected and *vice versa*.

Now consider a linear function of π's say

$$\sum_{1}^{k_1+k_2} a_i\tilde{\pi}_i,$$

a function involving parameters from both hypotheses. For finding a confidence interval for this note that

$$\text{Var.}\,(\tilde{Y}_i\, y) = \sigma^2 \text{ for all } i = 1, \ldots, k_1 + k_2$$

and

$$\text{Var.}\,\{\sum^{k_1+k_2} a_i\,(\tilde{Y}_i, y)\} = \Sigma\, a_i^2\,.\,\sigma^2 + \sum_{i\neq j=1}^{k_1+k_2} a_i a_j\ \text{Cov}\,\{(\tilde{Y}_i, y)\,(\tilde{Y}_j, y)\}$$

$$= \sigma^2\,[\Sigma\, a_i^2 + \sum_{i\neq j=1}^{k_1+k_2} a_i a_j\,(\tilde{Y}_i, \tilde{Y}_j)] = \sigma^2 . k \text{ say}$$

where $(\tilde{Y}_i\,.\,\tilde{Y}_j)$ is the scalar product. Then the confidence interval for

$$\sum_{1}^{k_1+k_2} a_i\tilde{\pi}_i$$

is given by

$$\Sigma\, a_i\,(\tilde{Y}_i, y) - \delta\,\sqrt{k} \leqslant \sum_{1}^{k_1+k_2} a_i\tilde{\pi}_i \leqslant \Sigma\, a_i\,(\tilde{Y}_i, y) + \delta\,\sqrt{k}\,* \tag{2.7}$$

If we confine only to contrasts either in $\pi_1, \ldots, \pi_{k_1}$ or $\pi_{k_1+1}, \ldots, \pi_{k_1+k_2}$ alone, the expression simplifies considerably giving the confidence interval for $\sum_{1}^{k_1} l_i\tilde{\pi}_i$ say, as

$$\Sigma\, l_i\,(\tilde{Y}_i\, y) - \delta\,\sqrt{\Sigma\, l_i^2} \leqslant \sum_{1}^{k_1} l_i\tilde{\pi}_i \leqslant \Sigma\, l_i\,(\tilde{Y}_i\, y) + \delta\,\sqrt{\Sigma\, l_i^2} \tag{2.8}$$

Method (*b*).—Applying Schwartz's inequality to the first half of relation (2.3) we get the confidence region and confidence intervals from

$$\sum_{1}^{k_1} [(U_n, y) - \phi_n]^2 \leqslant C_1. \tag{2.9}$$

* The result given by Ghosh (1955) is true only for the particular case (2.8) and does not work for the general situation as implied in that paper. It may be noted that much of the supposed simplicity of method (*a*) is taken away by this correction.

From the second half of relation (2.3) after applying Schwartz's inequality we get:

$$\sum_{1}^{k_2} [(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}]^2 \leqslant \sum_{1}^{k_1+k_3} [U_i, y) - \phi_i]^2. \dagger \tag{2.10}$$

Now we must find constants C_1 and C_2 such that the probability:

$$P_r\, [\sum_{1}^{k_1} \{(U_n, y) - \phi_n\}^2 \leqslant C_1, \sum_{1}^{k_1+k_2} \{(U_i, y) - \phi_i\}^2 \leqslant C_2 = 1 - a. \tag{2.11}$$

Instead of solving this integral Ghosh had suggested the use of the joint distribution of

$$\sum_{1}^{k_1} [(U_n, y) - \phi_n]^2 \text{ and } \sum_{k_1+1}^{k_1+k_2} [(U_n, y) - \phi_n]^2$$

in evaluating C_1 and C_2. But this gives only an upper bound of the simultaneous level of significance. Another approximation consists in using an inequality due to Kimball (1951) which is very simple for computations for C_1 and C_2 will then involve only the consultation of F-tables.

SECTION 3. RELATIVE MERITS

The comparison between different methods may be made either in terms of the volume of confidence regions of π's or in terms of lengths of confidence intervals for any standard function of π's. But, in fact, we need a comparison in parts, for Group I parameters alone and for Group II parameters alone but not both together; therefore the overall volume in $(k_1 + k_2)$ dimensions will not be suitable. It can be further seen by the known fact that area of a circle is smaller than that of a square of a side equal to the diameter.

We shall first consider the exact evaluation of (2.11). Putting

$$G_1 = \frac{\sum_{1}^{k_1} [(U_i, y) - \phi_i]^2}{n_e S_e^{\,2}}$$

† Ghosh (1955) had actually given an inequality more conservative than (2.9) by starting from $\gamma^{-1}\tilde{Y}_{II}$ and normalizing the k_2 individual components. As is clear, that transformation is totally unnecessary. In fact that inequality for second group of parameters is so conservative that method (*b*) would always be far inferior to method (*a*) specially for the contrasts of second group of parameters and the corresponding test of hypothesis.

and

$$G_2 = \frac{\sum_{1}^{k_2} [(U_{k_1+i}, y) - \phi_{k_1+i}]^2}{n_e S_e^2}$$

we have

$$P_r [G_1 \leqslant C_1,\ G_1 + G_2 \leqslant C_2] = 1 - a.$$

The joint distribution of G_1 and G_2 is,

$$C(k_1, k_2; ne) \frac{G_1^{(k_1-2)/2} G_2^{(k_2-2)/2}}{(1 + G_1 + G_2)^{(k_1+k_2+n_e)/2}}$$

so that transforming to $Z_1 = G_1$ and $Z_2 = G_1 + G_2$ and assuming limits $(0, \lambda_1)$ and $(0, \lambda_2)$ for the Z's we get

$$C(k_1, k_2;\ n_e) \int_0^{\lambda_1} \int_{Z_1}^{\lambda_2} \frac{Z_1^{(k_1-2)/2} (Z_2 - Z_1)^{(k_2-2)/2}}{(1 + Z_2)^{(k_1+k_2+n_e)/2}} dZ_2\, dZ_1 = 1 - a \quad (3.1)$$

Now, the lengths of confidence intervals for linear functions in $\pi_1, \ldots, \pi_{k_1}$ or $\pi_{k_1+1}, \ldots, \pi_{k_1+k_2}$ would be proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$ respectively. If the distinction between the first group of parameters and second is just arbitrary we can put $\lambda_1 = \lambda_2$, otherwise the ratio λ_1/λ_2 may be fixed at $k_1/k_1 + k_2$ or any other convenient quantity depending on the relative importance of the two groups or *a priori* knowledge.

If k_1 and k_2 both are even integers (3.1) can be simplified by expanding $(Z_2 - Z_1)^{(k_2-2)/2}$, etc., and we shall get a few incomplete B-integrals, so that λ_1 and λ_2 can be easily evaluated by using Pearson's Tables. But if $\lambda_1 = \lambda_2 = \lambda$ the method reduces to (*a*) and λ becomes simply

$$\frac{k_1 + k_2}{n_e} \cdot F_{1-a,\ k_1+k_2,\ n_e} = \lambda_a \text{ say.}$$

The advantage of this method (*b*) thus lies only in using two inequalities (for the two groups of parameters) instead of one. In general $\lambda_1 < \lambda_a$ and $\lambda_2 > \lambda_a$.

For comparing with method (*a*) we note that (2.5) gives the half length of the confidence intervals for a normalized linear function of $\pi_1, \ldots, \pi_{k_1}$ or $\pi_{k_1+1}, \ldots, \pi_{k_1+k_2}$ as δ_a where

$$\delta_a^2 = (k_1 + k_2) \cdot F_{1-a,\ k_1+k_2,\ n_e} \cdot S_e^2.$$

In (*b*), taking a normalized linear function of $\pi_1, \ldots, \pi_{k_1}$ say, we get the half-length as $\sqrt{\lambda_1 n_e S_e^2}$ so that λ_1/λ_a (or λ_2/λ_a) gives the ratio of (length)^2 and is tabulated below. For $n_e = \infty$ also, we are giving same values but here λ's are replaced by $n_e\lambda$'s.

TABLE I.

Values of λ_1/λ and λ_2/λ. $\alpha = \cdot 05$

Constants										
k_1 ..	2	2	2	2	4	4	2	2	4	4
k_2 ..	2	2	4	4	2	2	2	4	4	2
n_e ..	2	12	4	12	12	8	∞	∞	∞	∞
λ_a ..	38·5	1·081	9·21	1·500	1·90	2·68	9·49	12·59	15·51	12·59
λ_1 ..	24·0	0·713	3·81	0·697	1·15	2·11	6·25	6·19	9·93	9·87
λ_2 ..	48·0	1·426	10·43	2·091	2·30	3·17	12·50	18·57	19·87	14·81
λ_1/λ_a	0·62	0·66	0·41	0·46	0·61	0·79	0·66	0·49	0·64	0·78
λ_2/λ_a	1·24	1·32	1·24	1·38	1·21	1·18	1·32	1·47	1·28	1·17

It is clear from this that in many cases of practical interest λ_1 and λ_2 put together will give better results than method (*a*). There is a tendency that for smaller d.f.'s the advantage in using (*b*) may be considerable, specially when $k_2 > k_1$. Even for the case of $k_1 = k_2$ the method is generally better.

For $n_e = \infty$, the gain in using (*b*) shown above is the lower bound of the gain obtained. Actually the gain is more or less stabilized as n_e increases from 12 and so this value can be used as a lower bound. The computations for $n_e = \infty$ run on similar lines though they are a little simpler, being based on Incomplete γ-functions.

SECTION 4. METHOD (*d*)

We had the relation

$$\tilde{Y}_{II} = \beta\, U_I + \gamma\, U_{II}$$

which yielded

$$|\,(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}\,|^2 \leqslant \sum_{1}^{k_1+k_2} \{(U_j, y) - \phi_j\}^2 \text{ for all } i = 1, \ldots, k_2. \quad (4.1)$$

This inequality is conservative for we are not using the fact that rank $\beta \leqslant k_1$. In fact if $k_1 > k_2$, rank $\beta \leqslant k_2$ and so it is possible to get a better inequality.

Supposing then, $k_1 > k_2$ we have

$$\left.\begin{aligned} \tilde{Y}_{k_1+1} &= \beta_{11}U_1 + \ldots + \beta_{1k_1}U_{k_1} + \gamma_{11}U_{k_1+1} + \ldots \\ &\qquad + \gamma_{1,k_2} U_{k_1+k_2} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \tilde{Y}_{k_1+k_2} &= \beta_{k_2,1} U_1 + \ldots + \beta_{k_2,k_1} U_{k_1} + \gamma_{k_2,1} U_{k_1+1} + \ldots \\ &\qquad + \gamma_{k_2,k_2} U_{k_1+k_2} \end{aligned}\right\} . \qquad (4.2)$$

The k_2 vectors

$$\sum_1^{k_1} \beta_{ni} U_i, \; n = 1, \ldots, k_2$$

form a vector space whose rank $\leqslant k_2$. We transform $U_1, \ldots, U_{k_1}$ to another ortho-normal system $U_1', \ldots, U_{k_2}'$; $U_{k_2+1}', \ldots, U'_{k_1}$ such that the first k_2 vectors form a basis of the vector space of

$$\sum_1^{k_1} \beta_{ni} U_i \; (n = 1, \ldots, k_2).$$

This assumes that all rows of β are linearly independent and we can in general use 'r' as the number of vectors in the basis enumerated as $U_1', \ldots, U_r'$. As can be expressed as linear functions of $U_1', \ldots, U_r'$, equations (4.2) may be rewritten as:

$$\sum_1^{k_1} \beta_{ni} U_i$$

$$\left.\begin{aligned} \tilde{Y}_{k_1+1} &= \sum_1^{r} \beta_{1,j}' U_j' + \sum_1^{k_2} \gamma_{1,j} U_{k_1+j} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \tilde{Y}_{k_1+k_2} &= \sum_1^{r} \beta_{k_2,j}' U_j' + \sum_1^{k_2} \gamma_{k_2,j} U_{k_1+j}. \end{aligned}\right\} . \qquad (4.3)$$

Since all $U_1, \ldots, U'_r$; $U'_{k_1+1}, \ldots, U'_{k_1+k_2}$ are mutually orthogonal, each of unit length, the restrictions on β''s and γ's are:

$$\sum_{j=1}^{r} \beta_{ij}'^2 + \sum_{j=1}^{k_2} \gamma_{ij}^2 = 1 \text{ for all } i = 1, \ldots, k_2.$$

Applying Schwartz's inequality on relations (4.3) we get:

$$|(\tilde{Y}_{k_1+n}, y) - \tilde{\pi}_{k_1+n}|^2 \leqslant \sum_1^r [(U_i', y) - \phi_i']^2 + \sum_1^{k_2} [(U_{k_1+j}, y) - \phi_{k_1+j}]^2$$

$$\text{for all } n = 1, \ldots, k_2 \qquad (4.4)$$

where $\phi_i' = E\{(U_i', y)\}$ = a linear function of $\tilde{\pi}_{k_1+1}, \ldots, \tilde{\pi}_{k_1+k_2}$. For the first k_1 parameters, the inequality is the same as in method (*b*), *i.e.*,

$$|(\tilde{Y}_n, y) - \tilde{\pi}_n|^2 \leqslant \sum_1^{k_1} [(U_i, y) - \phi_i]^2 = \sum_1^{k_1} [(U_i', y) - \phi_i']^2$$

$$n = 1, \ldots, k_1 \qquad (4.5)$$

since sum of squares is in-variant under orthogonal transformation.

The optimum procedure in this case would depend upon the rank of β, however by putting $r = k_2$ we get an upper bound of level α.

Let

$$\sum_1^{k_2} [(U_i', y) - \phi_i']^2 = \chi_1^2, \quad \sum_{k_2+1}^{k_1} [(U_i', y) - \phi_i']^2 = \chi_1'^2$$

and

$$\sum_{j=k_1+1}^{k_1+k_2} [(U_j, y) - \phi_j]^2 = \chi_2^2,$$

then we shall use the joint distribution of

$$G_1 = \frac{\chi_1^2}{n_e S_e^2}, \quad G_2 = \frac{\chi_1'^2}{n_e S_e^2} \quad \text{and} \quad G_3 = \frac{\chi_2^2}{n_e S_e^2}$$

here χ_1^2, etc., are all independent χ^2-distributed variables.

We need two constants λ_1 and λ_2 such that

$$P_r [G_1 + G_2 < \lambda_1, \; G_1 + G_3 < \lambda_2] = 1 - \alpha. \qquad (4.6)$$

The lengths of $100(1-\alpha)\%$ confidence intervals for any normalized linear function of $\tilde{\pi}_1, \ldots, \tilde{\pi}_{k_1}$ or $\pi_{k_1+1}, \ldots, \tilde{\pi}_{k_1+k_2}$ are respectively proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$. Since the distinction between first group of parameters and second group may be arbitrary, we put $\lambda_1 = \lambda_2 = \lambda_d$ which simplifies both the computations and the comparison with method (*a*).

The joint distribution of G_1, G_2, G_3 is:

$$C\,(k_2, k_1 - k_2, k_2;\ n_e)\ \frac{G_1^{(k_2-2)/2}\, G_2^{(k_1-k_2-2)/2}\, G_3^{(k_2-2)/2}}{(1 + G_1 + G_2 + G_3)^{(k_1+k_2+n_e)/2}}$$

Transforming to $Z_1 = G_1$; $Z_2 = G_1 + G_2$ and $Z_3 = G_1 + G_3$, we get from (4.6):

$$C\,(k_2,\ k_1 - k_2, k_2;\ n_e) \int_0^{\lambda} \int_{z_1}^{\lambda} \int_{z_1}^{\lambda} \frac{Z_1^{(k_1-2)/2}\,(Z_2 - Z_1)^{(k_1-k_2-2)/2}\,(Z_3 - Z_1)^{(k_2-2)/2}}{(1 + Z_2 + Z_3 - Z_1)^{(k_1+k_2+n_e)/2}}$$

$$\times\, dZ_3\, dZ_2\, dZ_1 = 1 - \alpha \text{ for } Z_2 \geqslant Z_1,\ Z_3 \geqslant Z_1 \text{ always.} \qquad (4.7)$$

In general, the integral $I(\lambda, k_1, k_2)$ on the left-hand side of (4.7) can be evaluated by successive iteration and quadrature methods; but for k_1 and k_2 as even integers, the integrand can be expanded in powers of Z_2, Z_3 and Z_1 and each term can be further reduced on integration by parts to a double integral which, on integration by parts again, gives a few Incomplete B-functions. The computations are lengthy but for small values of k_1, k_2 which are most useful in practice, they can be easily done.

It should be noted that this method is some improvement over the methods (*a*) and (*b*) only when $k_1 > k_2$. When $k_1 = k_2$ this method fails and one has to fall back on (*a*) ar (*b*) unless something be known about the reduced rank of β. We give below a lemma to show that the value of λ satisfying (4.6) decreases with decrease in the rank of and attains the minimum value for $\beta \equiv 0$.

Lemma.—Let χ_1^2, χ_2^2 and χ_3^2 be 3 independently distributed χ^2-variables with respective d.f.'s r, $v_1 - r$ and v_2, then if r is allowed to vary, the probability of the event $[\chi_1^2 + \chi_2^2 < \lambda,\ \chi_1^2 + \chi_3^2 < \lambda]$ increases as 'r' decreases.

Proof.—Since any $\chi_{(n)}^2$ can be expressed as a sum of squares of 'n' standard normal variates, the event $[\chi_1^2 + \chi_2^2 < \lambda,\ \chi_1^2 + \chi_3^2 < \lambda]$ can be written as:

$$\left[\sum_1^r x_i^2 + \sum_{r+1}^{v_1} x_i^2 < \lambda,\ \sum_1^r x_i^2 + \sum_{v_1+1}^{v_1+v_2} x_i^2 < \lambda\right]$$

which implies the event:

$$\left[\sum_1^s x_i^2 + \sum_{s+1}^{v_1} x_i^2 < \lambda,\ \sum_1^s x_i^2 + \sum_{v_1+1}^{v_1+v_2} x_i^2 < \lambda\right] \text{ if } s < r.$$

The smallest value of r being zero, the maximum probability will hold for the event

$$\left[\sum_{1}^{v_1} x_i^2 < \lambda, \sum_{v_1+1}^{v_1+v_2} x_i^2 < \lambda.\right].$$

Hence the result.

Dividing each χ^2 by a non-zero prositive quantity, say S_e^2 the result still holds and if S_e^2 has a probability distribution the result will hold for χ_1^2/S_e^2, χ_2^2/S_e^2 and χ_3^2/S_e^2 on integrating the probabilities with respect to S_e^2 over its whole range.

Now, since

$$\frac{\sum_{1}^{r} [(U_i', y) - \phi_i']^2}{\sigma^2} = \chi_{(r)}^2$$

say,

$$\frac{\sum_{r+1}^{k_1} [(U_i', y) - \phi_i']^2}{\sigma^2} = \chi^2_{(k_1-r)}$$

and

$$\frac{\sum_{k_1+1}^{k_1+k_2} [(U_j, y) - \phi_j]^2}{\sigma^2} = \chi_{(k_2)}^2$$

have the required distributions, we have

$$P_r\,[\chi_{(r)}^2 + \chi^2_{(k_1-r)} < \lambda, \chi_{(r)}^2 + \chi_{(k_2)}^2 < \lambda]$$
$$\leqslant P_r\,[\chi_{(s)}^2 + \chi^2_{(k_1-s)} < \lambda, \chi_{(s)}^2 + \chi_{(k_2)}^2 < \lambda] \quad s < r$$

and on dividing by S_e^2/σ^2 and integrating over S_e^2 we get the same inequality. Then, corresponding to equation (4.6) we may say that

$$P_r\,[G_1' + G_2' < \lambda,\ G_1' + G_3 < \lambda]$$

increases as 'r' decreases; where dashes denote that the exact rank 'r' is used instead of k_2.

Table II gives comparative values of λ needed in method (*d*) when r is assumed to be k_2 as well as when it is minimum, *i.e.*, 0. In any practical situation the actual value lies between these 2 bounds. The 'λ' for orthogonal case ($r = 0$) has again been chosen to be the same for both groups of parameters for ease in computations and comparison, though there are reasons that it should better correspond in

TABLE II

Showing upper and lower bounds of λ_d as compared with λ_a

Constants													
k_1 ..	4	4	4	6	6	6	6	6	6	8	2	2	4
k_2 ..	2	2	2	2	2	2	4	4	4	4	2	2	4
n_e ..	4	8	12	6	8	12	8	12	20	16	2	12	12
λ_a ..	9·24	2·69	1·50	5·53	3·44	1·90	4·19	2·30	1·17	1·81	38·5	1·087	1·90
λ_d ..	7·77	2·26	1·28	4·53	2·83	1·57	3·59	1·97	1·02	1·44	{24·0 48·0	0·713 1·426	1·15 2·30
$\lambda_{orth.}$ ..	7·00	2·06	1·15	5·40	2·75	1·52	2·98	1·62	0·83	1·32	28·8	0·837	1·36
λ_a/λ_d ..	1·19	1·19	1·17	1·22	1·22	1·21	1·17	1·15	1·15	1·26	{1·60 0·80	1·50 0·75	1·65 0·83
$\lambda_a/\lambda_{orth.}$ ..	1·32	1·31	1·30	1·26	1·25	1·24	1·41	1·42	1·42	1·38	1·33	1·30	1·40

some way to the d.f.'s of G_1 and G_2. These computations are based on the integral

$$C(k_1, k_2;\ n_e) \int_0^{\lambda_1}\int_0^{\lambda_2} \frac{G_1^{(k_1-2)/2}\ G_2^{(k_2-2)/2}}{(1+G_1+G_2)^{(k_1+k_2+n_e)/2}}\ dG_2\, dG_1 = 1-\alpha \quad (4.8)$$

which simplifies to a few Incomplete B-integrals. In the special case $k_1 = k_2$ when method (d) fails we are also giving 2 values of λ from Table I along with the single λ-value from (4.8) for each $k_1 = k_2 = k$.

As in Table I we have taken k_1 and k_2 both as even integers. The values for odd k_1 or k_2 can be obtained by simple interpolation from Table II.*

There are some important points to note from the above observations. Firstly the advantage in using method (d) is nearly always small, coming to approximately 10% decrease in the lengths of confidence intervals. Secondly for rank $\beta = 0$, *i.e.*, orthogonal, the advantage in using method (d) as compared to (a) is considerable and is nearly unaffected by changes in error d.f.'s. Thus we can specify the approximate gain likely to accrue by using method (d). This is so even for $k_1 = k_2$ when method (d) coincides with (b).

* If k_2 is even but not k_2, the integrals (4.7) and (4.8) can still be exactly evaluated by first integrating out G_3 in (4.7) and G_2' in (4.8) and then solving the resulting double or single integral.

If n_e is large, say > 30 large sample approximation may be used for computing λ_d and this will be based on incomplete γ-integrals of Pearson. Finney (1941) has given some results in this connection for the case of orthogonal hypotheses.

If one is not willing to interpolate between pairs $(k_1 k_2)$ upper and lower bounds can still be obtained. If k_1 is odd but k_2 even, these are given by $\lambda(k_1+1, k_2)$ and $\lambda(k_1-1, k_2)$ respectively. Further another upper bound can be computed from the consideration that

$$P_r[G_{1(k_2)} + G_{2(k_1-k_2)} < \lambda,\ G_{1(k_2)} + G_{2(k_2)} < \lambda]$$
$$\geqslant P_r[G_{1(k_2)} + G_{2(k_1-k_2)} < \lambda,\ G_1'{}_{(k_2)} + G_{2(k_2)} < \lambda]$$
$$= P_r[G_{4(k_1)} < \lambda,\ G_{5(2k_2)} < \lambda]$$

where k_1, etc., in brackets ( ) shows d.f., G_1' denotes that it is independent of G_1 though identically distributed.

Tests of hypotheses, etc.—From (4.4) and (4.5) we get

$$|(\tilde{Y}_n, y) - \tilde{\pi}_n|^2 \leqslant (G_1 + G_2)\, n_e S_e^2 \qquad n = 1, \ldots, k_1$$

and

$$|(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}|^2 \leqslant (G_1 + G_3)\, n_e S_e^2 \qquad i = 1, \ldots, k_2$$

As before by considering a linear function of π's and maximizing over the linear coefficients, we get the two confidence regions:

$$\left.\begin{aligned} (\pi_1, \ldots, \pi_{k_1}) &: \sum_1^{k_1} [(\tilde{Y}_n, y) - \tilde{\pi}_n]^2 = (G_1 + G_2)\, n_e S_e^2 < C_1 \\ (\pi_{k_1+1}, \ldots, \pi_{k_1+k_2}) &: \sum_1^{k_2} [(\tilde{Y}_{k_1+i}, y) - \tilde{\pi}_{k_1+i}]^2 \leqslant (G_1 + G_3)\, n_e S_e^2 < C_2 \end{aligned}\right\}. \qquad (4.9)$$

For testing H_1 and H_2 we get the same two test statistics, *viz.*,

$$T_1 = \frac{\sum_1^{k_1} (\tilde{Y}_n, y)^2}{n_e S_e^2}$$

and

$$T_2 = \frac{\sum_1^{k_2} (\tilde{Y}_{k_1+i}, y)^2}{n_e S_e^2}$$

and the common critical limit is λ_α of Table II, etc. The tests are obviously quasi-independent and the simultaneous level of significance of the simultaneous test, *i.e.* $[T_1 > \lambda, T_2 > \lambda]$ is bounded above by 'α' though it will be nearer to α than in method (*a*) or (*b*). Table III of appendix gives a number of representative values for the upper bound of λ_α for $\alpha = \cdot 05$.

For more than two hypotheses the method can be easily generalised by the same approach. When $k_1 = k_2 = \ldots = k_s = 1$, we get a special case of this which may be useful in experiments where the main interest is in many single d.f. contrasts as for example factorial experiments.

As can be easily seen this method does not require any more computations than in method (*a*). The essential thing is to obtain the mutually orthogonal vectors $\tilde{Y}_1, \ldots, \tilde{Y}_{k_1}; \tilde{Y}_{k_1+1}, \ldots, \tilde{Y}_{k_1+k_2}$. This can be done by Gram-Schmidt method and it will be found that the necessary scalar products can be obtained by writing the normal equations of 2-way classification with unequal number of observations, etc. An example is given below illustrating the type of practical situations where these methods may be useful.

SECTION 5. EXAMPLE

The 8 treatments below refer to an experiment conducted at Rice Research Station, Tirur, in Randomized Block Design. The idea was to study the effect of direct application of phosphatic manure to paddy or through green manure—crop preceding the paddy crop. The treatments were:

(A) No manure.

(B) 45 lb./acre of P_2O_5 applied at the time of transplantation.

(C) Sunhemp grown without P_2O_5 but 45 lb./acre of P_2O_5.... transplantation.

(D) Sunhemp grown with 45 lb./acre of P_2O_5.

(E) Dhaincha grown without P_2O_5 but 45 lb./acre of P_2O_5... transplantion.

(F) Dhaincha grown with 45 lb./acre of P_2O_5.

(G) Sasbania grown without P_2O_5 but 45 lb./acre of P_2O_5.. transplantation.

(H) Sasbania grown with 45 lb./acre of P_2O_5.

There are obviously two natural groupings: one of A and B with direct application of P_2O_5; the other of C....H where phosphate is given through green manure. The experimenter's interest is confined to comparisons within each group (1 and 5 d.f.'s respectively) and one between the average of the two groups. Thus we have 3 groups of new parameters:

(1) $t_A - t_B$;

(2) 5 contrasts between C ..H; and

(3) 1 contrast of group comparisons: $3(t_A + t_B) - (t_C + \ldots + t_H)$, where $t_A, \ldots, t_H$ represent the effects of treatments A....H and carry a restriction $\Sigma t_A = 0$.

This is a case of 3 orthogonal hypotheses with respective d.f.'s as $k_1 = 5$, $k_2 = 1$, $k_3 = 1$. The analysis of variance model is

$$y_{(r)} = \mu + b_r + a_1 (t_A - t_B) + \sum_{i=2}^{6} a_i$$

(5 orthogonal contrasts in $t_C, \ldots, t_H$)

$$+ a_7 [\tfrac{1}{2}(t_A + t_B) - \tfrac{1}{6}(t_C + \ldots + t_H)] + e$$

e being the error term, independently normally distributed, μ being the general effect and $a_1, \ldots, a_7$ known constants taking suitable values according as a particular observation y comes from one treatment or another. The 5 orthogonal contrasts may be formed by the help of orthogonal polynomials though it would be more relevant to take them as $t_C - t_D$, $t_E - t_F$, $t_G - t_H$, $t_C + t_D - t_G - t_H$ and $t_C + t_D - 2t_E - 2t_F + t_G + t_H$.

Method (*a*) would have us using a F-test on 7 d.f. for testing the homogeneity of the 8 treatments and can provide us confidence intervals for any other contrast in $t_A, \ldots, t_H$ apart from those considered above. But method (*b*) would give us substantially low lengths for the contrasts (1) and (3) considered above which are more important for the experimenter. This would, of course, increase the length of confidence intervals for the parameters in group (2). In this case, however, the groups are orthogonal and $\beta = 0$ and thus method (*d*) will be very effective and will give smaller confidence intervals for all three groups, than method (*a*). Thus combining the groups (1) and (3) we get two groups [(1), (3)] and (2), with 2 and 5 parameters respectively, from the last row of Table II, the square of the lengths of confidence intervals by method (*d*) will be about 20% smaller than by method (*a*). To use the method (*d*) one has to find λ_d from Table III (for $d{\cdot}f$ 5, 2 and values of n_e given there), whereas in method (*a*) one uses the F-table with $d{\cdot}f$ 7 and n_e.

REFERENCES

1. Roy, S. N. and Bose, R. C. .. "Simultaneous confidence intervals estimation," *Ann. Math. Stat.*, 1953, **24**, 513–36.
2. Finney, D. J. .. "The joint distribution of variance ratios based on a common error mean square," *Ann. Eug.*, 1941, **11**, 136–40.
3. Ghosh, M. N. .. "Simultaneous tests of linear hypotheses," *Biometrika*, 1955, **42**, 441–48.
4. Kimball, A. W. .. "On dependent tests of significance in the analysis of variance," *Ann. Math. Stat.*, 1951, **22**, 600–02.
5. Nair, K. R. .. "The studentized of the extreme mean square test in the ANOVA," *Biometrika*, 1948, **35**, 16–31.
6. Pearson, K. .. *Tables of Incomple B-functions*, The Biometrika Office, London, 1934.
7. Ramachandran, K. V. .. "On simultaneous ANOVA test," *Ann. Math. Stat.*, 1956, **27**, 521–26.
8. ———— .. "Studentized largest and smallest χ^2" (Abstract), *Ibid.*, 1956, **27**, 860.
9. Scheffé, H. .. "A method of judging all contrasts in the ANOVA," *Biometrika*, 1953, **40**, 87-104.

APPENDIX

TABLE III

Showing values of λ_d and $n_e . \lambda_d$ (in brackets) $\alpha = \cdot 05$

Pairs (k_1, k_2) \ n_e	4	6	8	10	12	30
(3, 2)	6·93	3·25	2·06	1·485	1·156	0·378
	(27·72)	(19·50)	(16·48)	(14·85)	(13·87)	(11·34)
(4, 2)	7·77	3·60	2·26	1·630	1·275	0·411
	(31·08)	(21·60)	(18·08)	(16·30)	(15·30)	(12·33)
(5, 2)	8·85	4·03	2·53	1·816	1·411	0·442
	(35·40)	(24·18)	(20·24)	(18·16)	(16·93)	(13·26)
(6, 2)	9·87	4·53	2·83	2·040	1·574	0·498
	(39·48)	(27·18)	(22·64)	(20·40)	(18·84)	(14·94)
(7, 2)	11·1	5·11	3·16	2·268	1·737	0·571
	(44·4)	(30·66)	(25·28)	(22·68)	(20·84)	(17·13)
(8, 2)	12·4	5·66	3·50	2·501	1·929	0·611
	(49·6)	(33·96)	(28·00)	(25·01)	(23·15)	(18·33)
(5, 4)	12·3	5·62	3·49	2·489	1·924	0·616
	(49·2)	(33·72)	(27·92)	(24·89)	(23·09)	(18·48)
(6, 4)	12·9	5·82	3·59	2·561	1·975	0·625
	(51·6)	(34·92)	(28·72)	(25·61)	(23·70)	(18·75)
(7, 4)	13·7	6·08	3·75	2·666	2·066	0·646
	(54·8)	(36·48)	(30·00)	(26·66)	(24·79)	(19·38)
(8, 4)	14·1	6·33	3·94	2·804	2·160	0·674
	(56·4)	(37·98)	(31·52)	(28·04)	(25·92)	(20·22)
(8, 6)	17·8	8·14	5·01	3·541	2·705	0·844
	(71·2)	(48·84)	(40·08)	(35·41)	(32·46)	(25·32)

N.B.—For interpolating between two n_e-values for any pair (k_1, k_2) it is preferable to use $n_e . \lambda$ rather than λ-values.